Facilitating skill acquisition with video-based modeling worked examples

Inauguraldissertation
zur Erlangung des Doktorgrades der Philosophie
an der Ludwig-Maximilians-Universität München

vorgelegt von
Lena Zirn
aus Warstein
2015

Erstgutachter: Prof. Dr. Frank Fischer
Zweitgutachter: PD Dr. Karsten Stegmann
Datum der mündlichen Prüfung: 29.01.2015

Danke!

Diese Arbeit ist denjenigen gewidmet, die mich während der Doktorarbeit begleitet und unterstützt haben. Ohne die kollegiale und fachliche Unterstützung von Karsten Stegmann wäre die Arbeit in dieser Form nicht möglich gewesen. Hierfür ein riesengroßes Dankeschön! Frank Fischer hat mir beigebracht alle Aussagen kritisch zu hinterfragen, nie den Bezug zu Theorie und Erkenntnisgewinn aus den Augen zu verlieren und mich oft genug mit seiner Frage ‚Das verstehe ich nicht!' darauf hingewiesen, dass ich es selbst noch nicht ganz verstehe.
Meine vielen (Doktoranden-)Kollegen an der LMU waren seit Beginn der Arbeit meine Konstante: Julia, Freydis, Petra, Nicole, Maxi und ihr vielen anderen seid verantwortlich dafür, dass ich mich immer unterstützt und wohlgefühlt habe!

Nicht zu vergessen sind die vielen Studentinnen und Studenten, die ihre Abschlussarbeit in meinem Projekt geschrieben haben und mir in diesem Rahmen bei der Dateneingabe und Datenauswertung geholfen haben. Danke, ohne euch würde ich heute noch Videos kodieren!

Und *last but not least* möchte ich mich bei meinem Mann, meiner Familie und meinen Freunden bedanken! Danke, dass ihr mir immer versichert habt, wie gut ich alles schaffen kann, auch wenn ich es selbst nicht geglaubt habe! Danke für das Gegenlesen der Arbeit, Ablenkung in der freien Zeit und die Vorfreude auf viele gemeinsame Stunden nachdem die Arbeit fertig gestellt ist! Und obwohl er - absolut gesehen – wohl ein Grund für die verlängerte Dauer bis zur Fertigstellung dieser Arbeit war, möchte ich mich noch bei meinem Sohn bedanken. Dafür, dass du mir hilfst die Perspektive zu bewahren.

Table of Contents

Zusammenfassung

Unangenehme Gespräche, wie das Überbringen von schlechten Nachrichten, sind sowohl für den Arzt als auch für den Patienten belastend (Parle, Maguire, & Heaven, 1997; Rosenzweig, Reibel, Greeson, Brainard, & Hojat, 2003). Das sogenannte breaking bad news (BBN) ist eine aus mehreren Teilkompetenzen aufgebaute Fertigkeit und dadurch als komplex zu klassifizieren. Ein Teilaspekt ist das Eingehen auf die Emotionen des Patienten (Baile et al., 2000), was bislang nur unzureichend geschieht (Lesho et al., 2009). Um die Fertigkeit zu fördern können Handlungsprotokolle wie die SPIKES von Baile und Kollegen (2000) vermittelt werden. Allerdings fehlte in bisherigen Ansätzen oftmals ein theoretisches Modell hinter den Lehrinterventionen (Libert et al., 2001).

Um das Ziel einer Lehrintervention zum Erlernen einer komplexen Fertigkeit näher zu definieren, muss das Konzept ‚Fertigkeit' und deren Komponenten geklärt werden. Die Komplexität einer Fertigkeit ist immer in Relation zu der ausführenden Person und damit auch deren kognitiver Entwicklung zu setzen. Grundlegende Fertigkeiten sind einfache sensomotorische Handlungen, während sich komplexere Fertigkeiten aus mehreren Handlungskomponenten zusammensetzen (Fischer, 1980). Das gleiche Prinzip gilt für kognitive Fertigkeiten mit dem zusätzlichen Aspekt der Integration und Koordination der einzelnen Komponenten (Lim, Reiser, & Olina, 2009; Salvucci, 2013). Damit kann auch BBN als komplexe Fertigkeit kategorisiert werden (Meunier et al., 2013), zumindest für diejenigen, die noch keine Expertise in diesem Gebiet erworben haben.

Den Erwerb von kognitiven Fertigkeit beschreibt Taatgen (2013) in seinem Modell der *primitive elements theory* basierend auf der Annahme, dass Fertigkeitskomponenten miteinander zusammenhängen und sich überschneiden und jede Ausführung durch Regeln und wenn-dann Paradigmen enkodiert ist. Durch die Überschneidung der zugrunde liegenden Komponenten verschiedener Fertigkeiten findet Transfer statt (Taatgen, 2013). Bei der Förderung und Messung können komplexer Fertigkeiten einerseits in einzelne Komponenten zerlegt werden (part-task) oder aber mit ganzheitlicheren Lernaufgaben (whole-task) unterrichtet werden, wie dem 4C/ID-Modell von van Merriënboer, Clark, and de Crook (2002). In ihrer empirischen Studie konnten Lim und Kollegen (2009) positive Effekte eines modellierten whole-task Ansatzes feststellen. Im Vergleich zu einem part-task Ansatz, bei dem insgesamt 22 Einzel-Fertigkeiten demonstriert und reproduziert werden sollten, hatte die Bedingung in der zwei Beispiele mit Modell die auszuführende Fertigkeit zeigten positive Effekte auf die Ausführung ebendieser. Die positive Wirkung von modellierten whole-tasks kann teilweise auch auf die höhere Authentizität und dieser *real life problems* zurückgeführt werden, die den Lernen emotional involvieren (Anolli, Vescovo, Agliati, Mantovani, & Zurloni, 2006; Gulikers, 2006; Gulikers, Kester, Kirschner, & Bastiaens, 2008).

Eine wichtige Differenzierung bei der Analyse und Förderung von Fertigkeiten ist die Unterscheidung in kognitive und affektive Aspekte (Daniels et al., 2009; Op't Eynde & Turner, 2006). Kognitive Mechanismen sind die Grundlage komplexer Fertigkeiten und können

unterschieden werden in Gedächtnis, Wahrnehmung und Wissen die alle eine Rolle spielen beim Erwerb kognitiver Fertigkeiten (VanLehn, 1996). Vorwissen hat sich als entscheidender Faktor erwiesen indem neben Effekten auf den Trainingserfolg (Stark, Kopp, & Fischer, 2011) und Interaktionseffekten mit dem instruktionalen Ansatz (Atkinson, Derry, Renkl, & Wortham, 2000) Effekten auf die Wahrnehmung von Assessment (Gulikers et al., 2008) empirisch nachgewiesen werden konnten. Während die Konzeptualisierung von Emotion und Affekt noch nicht einheitlich ist, stimmen Lernwissenschaftler darin überein, dass beide Konzepte eine wichtige Rolle sowohl als Mediator im Lernprozess (Preston & de Waal, 2002) als auch in der Anwendung von erworbenen Fertigkeiten spielen (Tesser, Rosen, & Batchelor, 1972; Tesser, Rosen, & Tesser, 1971). Im Lernprozess konnten negative Effekte von Angst (Zeidner, 1998, 2007) und positive Effekte von Freude empirisch belegt werden (Daniels et al., 2009; Kim, Park, & Cozart, 2014). Lernaktivitäten und Lernkontext sollten demnach motivierend sein und Kontrollempfinden vermitteln um Angst im Lernprozess zu vermeiden. Die Wahrscheinlichkeit, dass eine erlernte Fertigkeit ausgeführt wird, wird maßgeblich vom affektiven Zustand des Lerners beeinflusst. So kann bis zu 30% der Varianz des Widerstreben schlechte Nachrichten zu überbringen, dem sogenannten MUM Effekt (Rosen & Tesser, 1970), durch Angst erklärt werden (Merker, Hanson, & Poston, 2010). Das tatsächliche Verhalten in emotional belastenden Situationen kann nach Anolli und Kollegen (2006) durch die vorherige Aneignung von Kommunikationsfertigkeiten positiv beeinflusst werden.

Bei der Messung von Fertigkeitserwerb müssen somit möglichst authentische Instrumente eingesetzt werden um reliabel und valide Aufschluss über Transfer in Problemlösesituationen des alltäglichen Lebens zu geben (Patrício, Julião, Fareleira, & Carneiro, 2013). Simulationen stellen eine Möglichkeit für realistisches Assessment dar, korrelieren in ihren Ergebnissen allerdings oft nur geringfügig mit traditionellen Wissenstests (Fischbeck, Mauch, Leschnik, Beutel, & Laubach, 2011).

Beispielbasiertes Lernen wurde empirisch als effektive Methode zur Förderung komplexen Fertigkeitserwerbs bestätigt. Zwei beispielbasierte instruktionale Ansätze, die von unterschiedlichen Forschungssträngen stammen, wurden von Renkl (2014) verglichen und in einem übergreifenden Rahmenmodell verbunden: Lösungsbeispiele (auch worked-examples genannt) und Lernen am Modell teilen verschiedene Aspekte und mehrere der zugrundeliegenden Lernmechanismen überlappen sich. Beispielsweise ist es bei beiden Ansätzen essentiell das effektivste Beispiel auszuwählen um den optimalen Lernerfolg zu erreichen. Die meisten empirischen Studien untersuchen bislang allerdings entweder worked examples oder Lernen am Modell.

Beim Lernen mit worked examples ist es essentiell, dass der Lerner Selbsterklärungen produziert (Chi, Bassok, Lewis, Reimann, & Glaser, 1989). Das gilt aber nur, wenn die Selbsterklärungen inhaltlich korrekt sind (Berthold & Renkl, 2009; Kuhn & Katz, 2009). Ein weiteres, häufig untersuchtes Konzept, ist der mentale Aufwand während der Bearbeitung von worked examples, auch cognitive load genannt (Sweller, 1994). Allerdings werden Definition und Operationalisierung kontrovers diskutiert. Fehlerhafte (erroneous) worked examples haben sich als effektive Variation erwiesen und es konnten positive Effekte auf die von den Lernenden gemachten Selbsterklärungen gezeigt werden (Siegler, 2002). Das lässt sich nach VanLehn (1999) mit einer Anregung zur Reflexion durch Sackgassen in Lösungsbeispielen erklären. Fehlerhafte Beispiele die von einem Modell präsentiert werden haben dagegen negative Effekte auf den Lernerfolg gezeigt (Bandura, 1977). Beim Lernen am Modell sind zentrale Aspekte empirischer Untersuchungen Charakteristika des Modells sowie die Beziehung von Beobachter und Modell. Letztere kann – wenn sie in einem mediierten Setting auftritt - mit dem Konzept der Parasozialen Interaktion (PSI) operationalisiert werden (Klimmt, Hartmann, & Schramm, 2006) Die Autoren unterscheiden verschiedene Aspekte dieser automatischen Rezipienten-Reaktion, nämlich affektive, kognitive und verhaltensbezogene PSI sowie den Aspekt der empfundenen Sympathie für das Modell. PSI muss allerdings von der Persönlichkeitseigenschaft Empathie abgegrenzt werden. Indem beide Ansätze beispielbasierten Lernens kombiniert werden, stellt sich die Frage nach dem Transfer der Befunde, beispielsweise der instruktionalen Unterstützung in Form von unvollständigen Beispielen und Selbsterklärungs-Prompts. Neben Inhalt und Struktur zeigt sich eine Tendenz hin zu Beispielen die video- und bildschirmbasiert sind. Damit wird eine Integration von menschlichen Modellen in worked examples möglich gemacht wird und die Beziehung vom beobachtenden Lerner zum Modell wird zu einem entscheidenden Faktor für den Lernerfolg.

Aus dieser empirischen und theoretischen Grundlage heraus ergeben sich Forschungsfragen, die mit zwei aufeinander aufbauenden Studien beantwortet werden sollen. Beide Datenerhebungen fanden in Kooperation mit der Chirurgischen Klinik der LMU München mit Medizinstudierenden des klinischen Studienabschnittes jeweils im Wintersemester statt.

Im Wintersemester 2009/10 wurden vier Fragestellungen untersucht. Die erste Fragestellung prüft den Effekt der Einführung eines videobasierten modellierten worked examples auf den Erwerb komplexer Fertigkeiten. Die Erwartung war, dass Lernen mit videobasierten modellierten worked examples im Vergleich zu textbasiertem Wissensinput einen positiven Effekt auf die Performanz der komplexen Tätigkeit hat. Mit der zweiten Fragestellung soll untersucht werden inwiefern die Selbsterklärungsunterstützung durch unvollständige videobasierte modellierte worked examples einen Effekt auf den Lernerfolg und den Verlauf der negativen Emotionen Angst und Hemmung hat. Daran anknüpfend fragt die dritte Fragestellung nach Effekten von Selbsterklärungsprompts in unvollständigen videobasierten modeliierten worked examples auf Lernen und Emotionen. Beide Ansätze zur Unterstützung von Selbsterklärungen bei den Lernenden sollten einen positiven Effekt auf die Aneignung der komplexen Fertigkeit des BBN haben. Im Rahmen einer weiteren Fragestellung sollen außerdem erwartete positive Zusammenhänge zwischen germaine cognitive load im Lernprozess mit Selbsterklärungsaktivität der Lernenden und dem Lernerfolg untersucht werden.

Während Fragestellungen zwei bis vier innerhalb der Stichprobe des Wintersemesters 2009/10 untersucht wurde, wurden für die erste Fragestellung die Fertigkeit des BBN der Studierenden aus dem Wintersemester 2009/10 mit der der Medizinstudierenden verglichen, die ein Jahr zuvor den gleichen Pflichtkurs ohne videobasierte modellierte Lösungsbeispiele besucht hatten. Alle Studierenden des Wintersemesters 2009/10 lernten mit einer online Lernumgebung, in der das videobasierte modellierte Lösungsbeispiel eingebettet war. BBN als Performanz der komplexen Fertigkeit wurde operationalisiert als Anzahl der SPIKES Schritte, die in einem Gespräch zwischen Lernenden und standardisiertem Patienten am Tag nach der Lerneinheit kodiert werden konnten.

Statistische Analysen zur Beantwortung der Fragestellungen bestätigen den erwarteten positiven Effekt der videobasierten modellierten Lösungsbeispiele auf den Fertigkeitserwerb. Dagegen haben unvollständige Beispiele zur Unterstützung von Selbsterklärungen einen negativen Effekt auf den Lernerfolg und Selbsterklärungs-Promts gar keinen messbaren Effekt. Die Werte von Angst und Hemmung bezüglich BBN sind nach dem Lösungsbeispiel geringer als bei der Vorbefragung, unterscheiden sich jedoch nicht zwischen den Bedingungen. Die erwarteten Zusammenhänge zwischen der Anzahl korrekter Selbsterklärungen und der Performanz der komplexen Fertigkeit konnten nicht nachgewiesen werden. Ebenso zeigten die Daten keine Zusammenhänge von Germane Cognitive Load und negativem Affekt mit dem Lernerfolg.

Die zweite Studie fand ein Jahr später im gleichen Setting statt und soll die Frage nach dem effektivsten Modell für den Erwerb komplexer Fertigkeiten beantworten sowie Erkenntnisse zum Prozess des Lernens mit videobasierten modellierten worked examples generieren.
Die erste Fragestellung untersucht den Effekt von Modellen in video-basierten worked examples die fehlerhaftes Verhalten zeigen und Rollenmodellen die sich im Status unterscheiden auf PSI, Cognitive Load und die Generierung von Selbsterklärungen. Die Erwartung war, dass eine stärkere Identifizierung mit einem studentischen Rollenmodell mit höheren Werten in PSI und mehr Selbsterklärungen einhergehen sollte im Vergleich zum Expertenrollenmodell. Weiterhin wurde erwartet, dass das fehlerhafte Modell stärkeren Extraneous Load bei den Lernenden hervorruft als das korrekte Modell. In der zweiten Forschungsfrage sollte der Effekt der unterschiedlichen Modelle auf den Lernerfolg verglichen werden. Dabei wurde erwartet, dass das studentische Modell mit korrektem Verhalten die besten Effekte auf die Performanz und den Wissenserwerb bzgl. der komplexen Fertigkeit haben sollte. In einem nächsten Schritt sollen in der dritten Forschungsfrage die Beziehungen von Persönlichkeitsmerkmalen des Lerners, Lernprozess und Fertigkeitserwerb analysiert werden. Positive Zusammenhänge wurden erwartet zwischen der Empathie des Lernenden und PSI sowie zwischen PSI, der empfundenen Sympathie für das Modell und den Selbsterklärungen mit dem Fertigkeitserwerb. Negative Zusammenhänge mit der Fertigkeit des BBN wurden dagegen erwartet für die Emotionen Angst und Hemmung des Lernenden sowie Extraneous Load. In einem letzten Schritt sollten zwei Modelle zur Varianzaufklärung beim Lernen mit videobasierten modellierten worked examples getestet werden.
Eine letzte Forschungsfrage soll den Effekt der Intervention durch video-basierte modellierte worked examples zum Erwerb komplexer Fertigkeiten im Medizinstudium klären. Die Hypothese war, dass es neben dem positiven Effekt der Einführung der Lösungsbeispiele

einen positiven Effekt der Weiterentwicklung des Lernmaterials von der ersten zur zweiten Studie gab.

Die Struktur der worked examples die in der zweiten Studie eingesetzt wurden ist im Gegensatz zur ersten Studie konstant. Die variierten Faktoren beziehen sich auf das dargebotene Modell und sind Status des Rollenmodells (hoher Status vs. gleicher Status) und Fehlerhaftigkeit des gezeigten Verhaltens (fehlerhaft vs. korrekt). Die Messinstrumente der ersten Studie wurden weitgehend übernommen. Leichte Adaptionen gab es bei der Messung des negativen Affekts und Cognitive Loads sowie in der Einführung des Konstrukts der Parasozialen Interaktion mit den Subskalen affektiver und kognitiver PSI und einem Item zur empfundenen Sympathie für das Modell.
Ergebnisse der statistischen Auswertungen zeigen Effekte der experimentellen Variation von Status auf affektive PSI mit höheren Werten für das Modell mit gleichem Status. Die Fehlerhaftigkeit des gezeigten Modell-Verhaltens hat einen Effekt auf die für das Modell empfundene Sympathie mit höheren Werten für das korrekte Verhalten. Dagegen wurden keine Effekte der Modell-Variation auf Selbsterklärungen und Cognitive Load gefunden. Bezüglich der zweiten Forschungsfrage nach Effekten auf den Lernerfolg zeigten sich entsprechend der Hypothesen Effekte der Fehlerhaftigkeit mit besserer Performanz nachdem mit korrekten Beispielen gelernt wurde und Effekte des Rollenmodells mit besserer Performanz nachdem mit dem Modell mit gleichem Status gelernt wurde. Allerdings konnten keine Effekte auf den Wissenserwerb bezüglich BBN gemessen werden. Analysen im Rahmen der dritten Forschungsfrage konnten den Rückgang von Angst bezüglich BBN der ersten Studie replizieren und zusätzlich zeigen, dass der Rückgang der Angst mit affektiver PSI im positiven Zusammenhang steht. Genau wie in der ersten Studie hat auch bei dieser Datenerhebung die experimentelle Variation des worked examples keinen Effekt auf die gemessene Angst der Lernenden. Das getestete Modell für den Faktor Fehlerhaftigkeit des gezeigten Verhaltens konnte bestätigt werden: Hemmung vor der Lerneinheit, Sympathie für das Modell, affektive PSI sowie extrinsischer Cognitiver Load konnten die Effekte des auf die Performanz der komplexen Fertigkeit erklären. Für den Effekt des Rollenmodells im worked example auf den Erwerb der komplexen Fertigkeit konnten keine mediierenden Faktoren gefunden werden, die größere Varianz aufklären können. Die Hypothese der letzten Forschungsfrage, nämlich dass die Weiterentwicklung des Lernmaterials nach der ersten Studie einen positiven Effekt auf den Erwerb der komplexen Fertigkeit hat, konnte bestätigt werden und der Gesamteffekt von der Kontrollbedingung der ersten Studie zur besten Bedingung der zweiten Studie ist mit einer Effektstärke von d = 1.91 als groß zu bewerten.

Aus den Ergebnissen der beiden Studien ergeben sich Schlussfolgerungen für die Mechanismen eines kombinierten instruktionalen Ansatzes von worked examples und modellierten Beispielen. So konnten mit unvollständigen modellierten worked examples nicht die erwarteten positiven Effekte auf die Anzahl der Selbsterklärungen und den Fertigkeitserwerb erzielt werden. Im Vergleich erzielten die Lerner bessere Ergebnisse, die mit vollständigen worked examples lernten. Die Anzahl von produzierten Selbsterklärungen während modellierter worked examples zeigte keine Zusammenhänge zum

Fertigkeitserwerb. Inhaltlich erwies sich das worked examples mit Übereinstimmung des Status von Rollenmodell und Lernenden, welches das korrekte Verhalten zeigt, als am besten geeignet um den Erwerb der komplexen Fertigkeit des BBN zu fördern. Korrektheit von Verhalten hat einen großen Effekt auf die empfundene Sympathie für das Modell, welche gemeinsam mit Parasozialer Interaktion mit dem Modell Varianz des Effekts auf den Fertigkeitserwerb aufklären kann. Extraneous load und Hemmung bezüglich BBN sind hinderliche Faktoren für den Fertigkeitswerwerb mit modellierten worked examples. Die Rolle von negativem Affekt beim Lernen mit modellierten worked examples konnte nicht gänzlich geklärt werden. Empathie der Lernenden hat lediglich einen kleinen Effekt auf den Fertigkeitserwerb und Angst wurde über alle Bedingungen reduziert. Beide stehen im Zusammenhang mit affektiver Parasozialer Interaktion. Alle Ergebnisse beziehen sich auf den Erwerb der komplexen Fertigkeit des BBN während keine Effekte auf den Wissenserwerb festgestellt werden konnten.

Limitierende Faktoren der Studien waren die zeitlichen Einschränkungen durch die Erhebung in einer Pflichtveranstaltung, welche das Design auf ein modelliertes worked example beschränkten. Des Weiteren sind beide Studien nicht in allen Belangen vergleichbar, da Skalen angepasst wurden und ein Informationstext von Englisch auf Deutsch übersetzt wurde. Die Bedingungen der zweiten Studie unterschieden sich neben den Faktoren Status und Fehlerhaftigkeit auch in der Person des gezeigten Modells. In der Erhebung der komplexen Fertigkeit wurden randomisiert verschiedene Patienten eingesetzt. Konfundierung dieses Aspekts kann nicht ausgeschlossen werden und limitiert die Aussagekraft der Ergebnisse.

Zukünftige Studien sollten den Aufbau der Studie um weitere Beispiele ergänzen, die alle das gleiche Modell in unterschiedlichen Bedingungen zeigen um ungewollte Effekte der Person des Modells auszuschließen. Die Lehreinheit sollte außerdem nach 30-34 Monaten wiederholt werden um einem Rückgang der erlernten Fertigkeit entgegen zu wirken. Um den positiven Effekt von vollständigen modellierten worked examples im Vergleich zu den unvollständigen Beispielen erklären zu können sollten Selbsterklärungen auch in der vollständigen Bedingung erhoben werden. Bezüglich der Erhebungsinstrumente sollte die Schwierigkeit des Wissenstest gesteigert werden und der Aspekt der Parasozialen Interaktion als interpersonalen Faktor zwischen Modell und Lernendem weiterhin erhoben werden. Die Hemmungsskala muss vor einem erneuten Einsatz validiert werden und die Standardisierten Patienten trainiert werden um objektive Rückmeldung geben zu können. Letztere könnten weiterhin durch kontrollierte Variation ihrer emotionaler Reaktion Einblicke in den Einfluss des Patienten auf BBN geben. Um weitere Einblicke in den negativen Effekt fehlerhafter modellierter worked examples zu bekommen, sollten Persönlichkeitsfaktoren der Lernenden erhoben werden. Eine
Generell ist bei ausreichender Stichprobengröße eine Auswertung mit Pfadanalyse zu empfehlen und eine Ausweitung auf andere Lernbereiche zu empfehlen. Eine Validierung des BNN Ratings durch Ärzte und Patienten wäre ebenfalls wünschenswert.

Bei der Gestaltung von Lerneinheiten mit modellierten worked examples zum Erwerb komplexer Fertigkeiten, die emotional negativ besetzt sind, sollte das Modell dementsprechend möglichst authentisch sein, als sympathisch empfunden werden und damit starke affektive Parasoziale Interaktion im Lernenden auslösen. Meiner Studien zufolge sind dies Modelle die korrektes Verhalten zeigen und dem Lerner ähneln.

Abschließend kann festgehalten werden, dass videobasierte modellierte worked examples eine geeignete Methode für den komplexen Fertigkeitserwerb von BBN in der Ausbildung von Medizinstudierenden sind. Die Erkenntnisse aus der Forschung zu worked examples und Lernen am Modell lassen sich nur teilweise übertragen. Dagegen hat es sich bestätigt den Aspekt der Parasozialen Interaktion in das Design mit aufzunehmen, da sie Varianz im Lernerfolg aufklären kann und Zusammenhänge zur Reduktion von Angst bei den Lernenden aufweist.

1 Introduction

'*The more costly and hazardous the possible mistakes, the heavier must be the reliance on observational learning from competent exemplars*' (Bandura, 1986, p. 20)

1.1 Problem Background

One area where peoples' fate and future is hanging in the balance is medical care in hospitals. If patients are not handled well by the doctor, e.g. by being provided with an adequate amount of information, they often develop mental health problems (Parle et al., 1997). Furthermore, the physicians themselves can be harmed by constant exposure to stress and thereby not able to reach their full potential in their jobs and private lives (Rosenzweig et al., 2003).

Breaking bad news (BBN), typical a task for physicians is defined by Ptacek and Eberhardt (1996) as a message that results in a cognitive, emotional, or behavioural deficit in the recipient. The activity of delivering such news is composed of a number of different skills which leads to the label of complex skill. It is not an easy task to perform and can be quite stressful for the doctor as well as the patient, in addition to the shock of receiving a severe diagnosis. Studies have shown that doctors often lack the ability to accurately estimate their patient's suffering (Lesho et al., 2009) and consequently do not react adequately. One of many negative consequences is the physician's tendency to provide systematically less information to elderly patients compared to younger patients (Gulbrandsen, Madsen, Benth, & Lærum, 2010) which is in contrast with the patient's need for assurance about their condition and prognosis. Baile and colleagues (2000) report that the response to the patient's emotions is what is regarded as most problematic in the scope of communication situations - also by the physicians themselves!

To avoid those negative consequences medical students and practicing physicians need to be trained in complex communication-skills and perform the critical behaviour before they enter into a real life scenario. This has shown to lead to more compliance and satisfaction on the patient's side and reduces burnout in physicians (Anolli, Vescovo, Agliati, Mantovani, & Zurloni, 2006; Fallowfield & Jenkins, 2004). One practical solution may be the introduction of steps for how to communicate with patients in difficult situations - like the SPIKES protocol for BBN (Baile et al., 2000). Ptacek, Ptacek, and Ellison (2001) report that some of those defined steps are being applied more often than others. Such observations can hint at what makes those communication situations stressful for physicians (McKee & Ptacek, 2001; Ptacek & McIntosh, 2009) and which steps are automatically applied in a physician's daily routine. There are a lot of articles on communication training for doctors. A major flaw is thus the consistent lack of theoretical models in the background of learning interventions (Libert et al., 2001) and the insufficient measurement of improvement of the doctor's quality of life - which is one of the overall goals of medical communication training (Libert et al., 2001).

Because the costs of one-trial learning by placing the future patients of those students at risk are too high medical students need to be trained in a safe environment with Standardised Patients before meeting real patients. In their review on communication training in oncology Libert and colleagues (2001) claim effective training should be interactive as well as dynamic and provoking a lot of effort from the participants. To control the problematic aspect of limited time, modelling through simulation is an efficient alternative for acquiring communication skills for difficult situations and can considerably shorten the learning process (Bandura, 1965; Lane, Slavin, & Ziv, 2001). When teaching a skill, worked examples can be used to structure the problem and effectively convey the steps of a possible solution. So far, worked examples are mostly applied in structured domains like mathematics, physics and computer programming, however researchers such as Rourke and Sweller (2009) provided the first hints that complex tasks can also be taught implementing this method. Rummel, Spada, and Hauser (2009) ask for an extension of worked examples in research to less structured domains because of a lack of existing research at present. Shen and Tsai (2009) summarised eight design principles for effective worked examples and while these guidelines are primarily based on traditional worked example research in domains with structured problems and clear and limited solution steps, most of the principles can be applied to more complex worked examples as well.

During training it is important to focus on individual characteristics of the learners as well as the Standardised Patients because the context – especially the social context – cannot be disregarded (Salomon, 1992). Until today, direct comparison of different patients[1] that focus on effects of their characteristics on the learning outcome is still missing (Cook, Erwin, & Triola, 2010).

The competence to perform complex skills such as communication in difficult situations needs to be trained. In order to define the concept of skills and its components the first chapter 2.1 of this work is dedicated to a definition of skill. Within this definition, I will attempt to differentiate in respect to the complexity of skills and report empirical results on how skill acquisition can be implemented effectively. Furthermore, authentic assessment will be defined and described as an appropriate means of measuring the effects of methods that foster skill acquisition.

Going deeper into the composition of skills and skilled behaviour, I will further distinguish between cognitive and affective aspects that are both relevant dimensions and need to be approached to support learning activities (Daniels et al., 2009; Op't Eynde & Turner, 2006).

Cognitive requirements in the form of knowledge, and motivational requirements in the form of motives and emotional dispositions (e.g. volition) are important prerequisites for competent behaviour (Kaufhold, 2006). Chapter 2.1.1 is therefore focused on cognitive skill acquisition and the assessment thereof. The subsequent chapter 2.1.2 introduces a

[1] Cook revised studies on health professions training with virtual patients.

definition of affect and emotions to subsequently present empirical findings on the influence of affect on the acquisition as well as the application of skills. Simulation is described as a mean to address negative emotions and avoid inhibition to perform a previously acquired skill.

In chapter 2.2 example-based learning is introduced as a method that has been empirically proven to be effective in fostering the acquisition of complex skills. Two instructional approaches are included that come from different lines of research but have been linked and compared resulting in a shared framework by Renkl (2014): Worked examples and modelling share many aspects and several learning mechanisms behind those two instructional approaches overlap, for instance the selection of the most effective example as a crucial factor for optimal learning outcomes. Still, most research is carried out with either worked examples (2.2.1) or modelling examples (2.2.2), and it is for this reason two distinct chapters display empirical results and theoretical frameworks, respectively. An aspect mostly explored for learning with worked examples is the role of self-explanation during the learning process and its effects on the learning outcome (2.2.1.2). Cognitive load (2.2.1.1) is a concept that is heavily intertwined with worked example research while the aspect of errors in the example (2.2.1.3) has been researched under different names for both methods. The relation of observer and model presented in the example is shown alongside characteristics of the model expected to be affected by the empathy of the learner (2.2.2.1). Parasocial interaction as a concept that has been developed to measure the relation of an observer to a model in a mediated setting is introduced in a sub-chapter and distinguished from the personality trait empathy (2.2.2.2). By combining both approaches to modeling worked examples, the question of transferability from one approach to the other arises and is discussed throughout the chapter on example-based learning. Especially in terms of instructional support the main body of research stems from empirical studies with worked examples but is not necessarily limited to this instructional approach (2.2.3).

Besides content and structure of the learning input, there is a tendency towards on-screen examples based on videos. Chapter 2.3 focuses on new aspects that are relevant for video-based examples. Human models integrated in worked examples and displayed on screen make the relationship of the learner to the model a crucial factor for the efficiency for learning outcomes.

In a concluding chapter a synthesis for complex skill acquisition is drawn (2.4) based on the findings from the previous chapters. The final chapter prior to the research questions briefly describes the approach of design-based research (2.4.1) that aims at directly combining research and practice in the field of learning by, inter alia, iteratively developed learning interventions.

This work includes two consecutive studies. In chapter 3 research questions to be targeted in a first round of data collection are presented and hypotheses for each of the questions are illustrated. The 1st study aims at exploring the most effective structure for video-based modeling worked examples by transferring findings from worked example research to modeling worked examples for complex skill acquisition in the setting of medical education.

Following the description of participants, procedure and learning environment for this study (4), the statistical results are presented (5) and linked to the hypotheses in a conclusion on the results (5.3).
A following study in the same setting aims at answering research questions on the most effective model and further insight in the process of learning with video-based modeling examples (6). The structure of the examples is constant between the conditions and the factors varied are the role-model and errors in the example. In addition, both rounds of data collection are analysed in respect to comparability and effect on enhancing complex skill acquisition.
After describing the sample as well as new instruments used during the 2nd study (7), empirical data is presented for the results and the findings are linked to the research questions (8).

In the final discussion the main outcomes are presented and summarised (9.1.). Subsequently those results are narrowed by discussions on limiting aspects and offer advice on research questions to be investigated in the future (9.2).

2 Skill Acquisition through Modeling Worked Examples

In order to understand how modeling worked examples can foster the acquisition of skill it must first be understood what a skill is and how it can be acquired. At first the concept of skill is described by presenting and defining the cognitive and affective aspects of it. Furthermore, different models of skill acquisition are outlined that have been empirically tested and should lead to a differentiated understanding of the underlying mechanisms. Subsequently, the instructional approach of worked examples is introduced as a method to foster skill acquisition. Crucial aspects of learning with worked examples that will be elaborated in more depth are self-explaining during learning and how to foster it. Moreover, the approach of modelling which can be allocated in the field of Bandura's social-cognitive approach to learning will be described and linked to learning with worked examples. A mechanism that is mostly accepted to lay behind the effects of learning with examples is the elicitation of different kinds of cognitive load that will be described before illustrating the special case of erroneous examples and under which circumstances they are effective. Finally, recent developments to present worked examples on screen in the form of videos will be introduced. When those video-based examples include a modeling case, concepts that look at the relation of the learner to the model become relevant which is why a chapter on empathy and parasocial interaction complete the theoretical foundation.

2.1 Skills

The oxford dictionary defines skill as 'the ability to do something well'[2]. The concept is thus clearly distinguished from knowledge by adding the aspect of performance. Besides being able to put knowledge into practice it is furthermore crucial for skilled behaviour to react to the environmental affordances (Fischer, 1980). Skill is thus always determined by the organism itself with its mental models as well as the environment requiring appropriate behaviour of the person.

When reviewing the literature on skill, there are different kinds of skills addressed: social skills, manual skills, cognitive skills, academic skills, language, and perceptual-motor skills[3]. Fischer (1980) developed a theory that is based on the hierarchy of skills in order to predict development of cognition. Cognitive development is thereby characterised as an increase in skill complexity starting with sensory-motor actions and leading to representations and finally abstractions.

Labelling skills as being complex is foremost prevalent in the context of motor skills. There, the complexity rises with the number of components a movement includes. But also domains such as hospital scenarios including patients and operations are often categorised as complex. In the domain of cognitive skills, most definitions allocate a skill as complex if it

---

[2] http://www.oxforddictionaries.com/definition/english/skill

[3] This list of skills is not exhaustive and just an insight in the different areas of skills mentioned in the literature described.

consists of multiple skills that require integration and coordination (Lim et al., 2009). Meunier and colleagues (2013) base their categorisation of complexity on the number of relevant dimensions that have to be dealt with during the activity and thereby label the skill of BBN as complex. This is a similar approach to the one applied by van Zundert, Sluijsmans, Könings, and Van Merriënboer (2012) in their empirical study on peer assessment skills. Both research groups rather describe the complexity of the task than the complexity of the skill. In the latter paper, skills are furthermore categorised as domain specific vs. 'twenty-first century skills' (Scardamalia, 2001). Kneebone and colleagues (2006) also focus on the complexity of the setting the performance (*here:* BBN) is located in: Low-complexity procedures like closure of a wound, mid-complexity procedures that require more sophisticated skills as well as adequate handling of patients, and high-complexity procedures that include every step from meeting a patient, diagnosing him and decide for an operation. Here, the number of skills that need to be integrated in order to successfully manage a task is taken into account. The integration and reuse of skills is also the basis of the model of cognitive skill acquisition proposed by Salvucci (2013). Single skills are labelled component skills that can be combined with other component skills to more complex skills. As even everyday skills such as shaking hands can be decomposed into component skills the categorisation of a skill as complex is always relative to the ability of the performing person. Experts might classify skills of their field of expertise as relatively simple while the same procedures are highly complex for novices.

Besides differences in the definition of complexity, most skills are composed of more than one element. To learn how to perform a skill, different models of skill acquisition have been developed. Different authors refer to this concept in altered terms. While VanLehn (2006) decomposed knowledge into independent knowledge components, Taatgen's (2013) model of cognitive skill acquisition is based on the assumption that skills are interrelated and overlapping and is called the *primitive elements theory*. The knowledge needed for the performance of a skill is encoded in rules or condition-action paradigms which is broken down to task specific and general elements (= primitive elements). The larger part is general and not task specific and can be combined to a multitude of different production rules. This can also explain the interrelatedness of skills that build on each other as well as the ability to transfer skills from one context to another that are similar in structure but share no content. By breaking down the skills into much smaller elements than it has been done before[4], more transfer can be explained. Transfer is defined as the 'overlap in combinations of PRIMs among skills' (p.26) with PRIMs being the smallest procedural elements Taatgen (2013) proposes: primitive information processing elements. Examples for PRIMs are *checking for visual input* and *copying visual input into retrieval*. The limited number of PRIMs is controlled by declarative operators that consist of linked memory items. Sources of operators can be instruction from external sources or reasoning. The more skilled a learner is, the more all necessary PRIMS are combined to productions that incorporate all elements to one step.

[4] A typical entity that has been described as being the basis for transfer are principles that may include a formula, the meaning of the symbols included and the area of application (VanLehn, 1996).

Traditionally, complex tasks have been broken down into smaller units when being taught. Following critique on this approach and longing for more authentic learning tasks, Lim and colleagues (2009) investigated the effect of a whole task vs. a part task. In one condition pre-service teachers (undergraduate students) had to perform the complex cognitive skill of preparing a grade book by using Excel (whole-task condition) while in the other condition the task was split into smaller tasks (isolated part-task condition). The instructional approach applied in the whole task condition is the 4C/ID-Model by van Merriënboer and colleagues (2002) that combines four components (learning (whole-) task, supportive information, procedural information, and part-task practice). Even though the model includes part-tasks, the key emphasis is on the whole task and operationalised here as modelling example. In addition to the task format, Lim and colleagues (2009) also tested for effects of learners' high vs. low prior knowledge. A knowledge pre-test measured 16 relevant skills. 29 learners were classified with low-prior knowledge while 26 reached high-prior knowledge values. Dependent variables were part-task achievement, attitudes, and time on task.

In the part-task condition 22 skills were demonstrated and had to be reproduced by the learners. Only after all skills were learned, the whole-task had to be performed.

In the whole-task condition two modeling examples were provided that showed the behaviour the learners adjacently had to apply to the same type of task. Correspondent to the 4C/ID approach, the learners were furthermore assisted with supportive and procedural information.

Skill acquisition as dependent variable was measured with a part-task and a whole-task test two days after the training. Other dependent variables were transfer (application of Excel in another case), time on test, and the learners' attitudes (questionnaire on motivational reaction to instructional material with four subscales attention, relevance, confidence, and satisfaction).

Results showed that there is no effect of instructional approach on the skill acquisition measured with part-tasks. When measuring with whole-task, there was a significant medium effect of instructional approach with those who learned with whole-task scoring better. No effect of prior knowledge on skill acquisition and no interaction effects with instructional approach could be reported. In the transfer test, the positive effect of the whole-task instructional approach was even larger and high prior knowledge had a significant positive effect. There was no interaction of instruction and prior knowledge on the transfer test. The learners' attitude was only influenced by prior knowledge on the level of confidence with those who had high prior knowledge indicating to be more confident.

This study thus supports the hypothesis that learning with a whole-task approach prepares better for later performance (whole-task as well as transfer) than instruction on the basis of isolated part-tasks. This finding is independent from the learners' prior knowledge. The authors note that the results may be restricted to relatively easy tasks as the scores in the post-test were rather high (score of 80% and 89% respectively).

In general, generic skill development and deep learning are positively affected by the authenticity of assessment form as well as of the task (Gulikers, 2006). Relatively complex tasks like role-play are well received by students. Though, a limiting factor is the perception as authentic. A study by Gulikers and colleagues (2008) has shown that the perception of assessment through role-play as authentic has a positive effect on the perception of generic skill development. The construction of a learning intervention as realistic and authentic is also a way to ensure transfer in later performance and not construct in inert knowledge that is not implemented in real life situations (Renkl, Mandl, & Gruber, 1996). This can be done by including real world problems (Lim et al., 2009), construct emotionally involving learning material (Anolli et al., 2006), and allow the learners to integrate all skills that are needed for the real task (Van Merriënboer & de Croock, 1992).

It can thus be concluded that the instructional approach for skill acquisition should be based on authentic whole-tasks combined with some support and vary in difficulty when complex skills are the learning goal (Lim et al., 2009). It is still an open issue whether this is also true for tasks that are experienced as difficult by the learners. Generally, to classify the complexity of a skill the distance between the behaviour the learner is already capable of and the targeted skill must be taken into account.

2.1.1 Cognitive Aspects of Skills

After the introduction of different conceptualisations of complex skills, different approaches to skill acquisition and the assessment thereof will be described in order to identify effective methods as well as crucial moderators and mediating factors.
Cognitive mechanisms are the core basis of any complex skill. Examples are memory, perception and knowledge. The latter knowledge component has been investigated most. It plays a major role in the acquisition of skills and the achievement of expertise in a performance. The level of relevant prior domain knowledge that is available to build on constitutes a factor that needs to be taken into account in the construction of learning material as well as the selection of methods and assessment. In his review on cognitive skill acquisition, VanLehn (1996) discriminates three phases that are not totally distinct. In the first phase knowledge relevant for the domain is integrated in existing cognitive structures. The second phase is labelled *intermediate phase* and consists of the application of domain knowledge to problem solving which is brought to perfection in the third phase. Problem-solving can either consist of one or more principles that have to be applied with an extension of learning time the more principles are involved. But before the application phase, it has to be retrieved from memory which is often supported by hints. Following the retrieval, the problem at hand has to be connected or mapped with the principles, not only by looking for surface matches, but by taking into account the underlying structure. This step is followed by the application of the principle and finally generalisation to other problems that are structured equivalent.

Some methods for knowledge acquisition, like worked examples are only effective when the level of prior knowledge is relatively low (Atkinson et al., 2000). On the other hand, some knowledge base is needed to effectively make use of the learning material. Rittle-Johnson, Star, and Durkin (2009) provided 236 8^{th} grade students in pairs with worked examples of multiple solution methods, multiple problem types, or multiple examples of one algebraic problem case. In all conditions, the worked examples were combined with reflection prompts. The intervention section was followed by four independent practice problems the students were asked to solve. The results showed that prior knowledge operationalized as use of algebra in a pre-test (yes vs. no) had an effect on which instructional condition worked best for the students. For those with low prior knowledge, comparing two identical solutions or sequentially studying had a positive effect on the post-test while comparing different methods had a negative effect. Students with high prior knowledge were able to benefit from comparing methods. This study illustrates the importance of prior knowledge as a prerequisite for positive effects of learning material that requires higher order cognitive processing.

Besides interaction effects of knowledge level and instructional approach, prior knowledge also has been shown to have even larger effects than any kind of intervention applied (here: worked examples and feedback, see 2.2.1.3) (Stark et al., 2011) and also has an effect on the perception of assessment. Gulikers and colleagues (2008) proved that students who perceive a role-play as authentic, have higher levels of perceived generic skill development by engaging in deeper learning. In their study senior students (skilled) perceive the physical context during the assessment within an authentic setting (role play) as less authentic compared to freshman students (little practical experience). The authors of the study discuss that explicit analytic steps are no longer needed by the experienced students and might even be demotivating when experience in performing the task has already been gained and led to a feeling of being already skilled. On the other hand relevant prior knowledge primarily needs to be activated in order to initiate a skilled performance.

There is also the case that learners with advanced levels of prior knowledge even suffer from too much instructional support. This effect has been labelled the expertise reversal effect (Kalyuga, Ayres, Chandler, & Sweller, 2003) and has been found in many areas of academic education (Kalyuga, Rikers, & Paas, 2012), see also the chapter on cognitive load 2.2.1.1.

In the context of medical students the knowledge that is acquired in university consists of three separate unities[5]. Specialised knowledge (=declarative knowledge) about the human physic, its functions and processes constitute the basis. But also motor skills (=application-oriented knowledge) need to be acquired. Examples for motor skills are physical examinations of parts of the body and later on methods to perform surgery. The third area of competence medical students need specialised knowledge for is physician-patient communication competence. This procedural skill has been formalised in protocols or schemas that seek to cover distinct situations.

It can thus be concluded that the method applied for cognitive skill acquisition as well as the assessment of learning should be adequate for the learners' level of prior knowledge. In their recent article on educational implications, Kalyuga and colleagues (2012) encouraged

[5] The categorisation of physicians' skills differ between publications. In their review on simulation in medical education, Lane et al. (2001) differentiate patient-centered, process-centered, and environment-centered skills.

research in medical education that experimentally varies case descriptions or the expertise of the learners.

Besides cognitive skill components, affective aspects are very often part of a skill as well and might call for changes empirical findings that are only related to cognitive skills. In any case a person's affective state is a crucial aspect in the acquisition of a new skill as well as in the application of it. Therefore, the next chapter introduces relevant affective aspects for skill acquisition and transfer to real world problems.

2.1.2 Affective Aspects of Skills

Emotions are allocated a major role in the context of learning. Independent from the theory, learning scientists agree that emotions play an important role during the learning process (Preston & de Waal, 2002) but also in the application of acquired skill (Tesser et al., 1972; Tesser et al., 1971). Recent research has underlined the importance to consider the strong interrelation of emotions with motivation. Academic emotions and motivation influence each other in both ways and have an effect on the action following: Emotions like anxiety can affect motivational matters like the desire to join a dance class that lead to absence in the lessons. But also the motivation to join a dance class can influence the emotional state resulting in anxiety while signing in for the course. Still, besides sometimes resulting in the same activity, motivation and emotions are not the same (Kim, 2012).

At first, affect and emotions have to be conceptualised as both concepts are often not well distinguished and the authors have not yet exclusively agreed upon the terminology. Emotions like anxiety are processes composed of many elements from the field of cognition, behaviour, physiology and affect according to Scherer (2000). Daniels and colleagues (2009) on the contrary state that the combination of multiple emotions of the same valence is conceptualised as affect and can either be positive or negative. Op't Eynde and Turner (2006) go so far to abandon the concept of cognitions and emotions as two distinct entities and propose a dynamic model for academic learning where emotions are composed of mutually regulating aspects like motivation, feelings, cognition, neurophysiology, and motor expressions. Both approaches share the conclusion that all aspects of learning need to be approached to support learning activities. In general, emotions can be assessed via questionnaires, via physiological attributes (e.g. skin resistance, heart rate, tonus of the muscles) or via analysing speech and facial expression. When measuring physiological attributes only the intensity and not the quality of the emotions is measured (Schachter, Singer, & Parrott, 2001).

2.1.2.1 Emotions in the Learning Process

Emotions play a crucial role in the acquisition of new skills and often can be considered as moderating or mediating the learning process. The most prominent relation of emotion and skill acquisition is the negative effect of test anxiety on achievement, as documented in longitudinal, as well as experimental studies by Zeidner (1998, 2007). Positive relations on the contrary have for instance been reported for the emotion enjoyment (Daniels et al., 2009; Kim et al., 2014).

In an experimental study, Daniels and colleagues (2009) showed, that emotions mediate the predictive relationship between a learner's mastery goal and the achievement following. 669 psychology students completed a questionnaire on initial affective experience in the beginning of their studies that were then categorised as hopeful or helpless. The authors have the hypothesis, that this affective state has an influence on performance and mastery goals that were assessed at the same time with four items of a questionnaire, respectively. Three discrete emotions enjoyment, boredom, and anxiety were assessed four month later with respectively six items taken from the AEQ – Achievement Emotions Questionnaire (Pekrun, Götz, & Perry, 2005). At the end of the introductory courses that lasted two semesters, the final grade that included five to six multiple choice tests was included as academic achievement. Every emotion was tested in a separate model to confirm the premise that emotions can predict achievement and mediate the effects goals have on achievement.

Results showed that the affective state of feeling helpless was negatively related to mastery goals while hopefulness positively predicted mastery and performance goals. Mastery goals were shown to predict each of the three emotions assessed (positive effect on enjoyment and negative effect on boredom and anxiety) that in turn significantly predicted achievement. Performance goals were positively related to anxiety. All three models were confirmed by the analyses, explaining between 21% and 24% of variance in achievement. There are no differences between male and female participants.

The authors conclude that it is possible to increase mastery goal adoption by reducing feelings of helplessness and increase feelings of hopefulness. Nevertheless, the authors assume that there are more influencing factors that mediate the models tested like learning strategies and student's self-regulation to be investigated in further studies.

In the setting of a non-collaborative online mathematic course, Kim, Park, and Cozart (2014) found that achievement emotions[6] constitute even 37% of variance in the student achievement (final grade scores), compared to only 13% accounted for by motivation and no additional effect of self-efficacy and cognitive strategy use. Among all emotions assessed, anger was the strongest individual predictor. As limitating factor to the external validity of the study, the authors name the developmental phase of the students (mean age = 16), namely puberty to be a possible reason for a higher occurrence of anger as well as little number of participants. Furthermore, the interaction between factors of the physical environment and learners' characteristics need to be taken into account as demonstrated in a study where the learners' prior level of anxiety had an interaction effect with a teacher- vs. learner-centred environment on performance (Dowaliby & Schumer, 1973).

Those results lead to the conclusion that emotions, especially anxiety must be considered when a learning environment is constructed (Kraiger, Ford, & Salas, 1993). Emotions can also be linked to specific operations, leading to reduced motivation and probability of

[6] Mathematic achievement emotions were boredom, anxiety, enjoyment, anger, shame, pride, and hopelessness.

performance (Deci & Ryan, 1985). The resulting affective and motivational state is labelled inhibition and investigated in the scope of medical education (Siebeck et al., 2011). Learning activities and the learning environment should thus be motivating as well as convey a sense of control to avoid anxiety during learning.

2.1.2.2 Emotions as Learning Outcome

Once a skill has been acquired the probability that it is applied has to be put centre stage. This approach has already been described by Bandura (1986) who underlined that the performance of acquired skills does depend heavily on motivational and affective states of the individual and that some skills are generally linked to negative emotions that make their performance less likely. Those 'affective outcomes' must be targeted within the learning goals to be able to impact on the probability that the skills taught are being applied (Parle et al., 1997).

The reduced probability to communicate negative information in contrast to positive information (keeping mum about bad news) independent from skill level has been introduced by Rosen and Tesser (1970) as the MUM Effect. They were the first to experimentally rule out other factors like rewards and punishment and the relation between communicator and recipient as moderating the effect of the reluctance. There was a significant effect of the value (positive vs. negative) of an overheard message on the transmission to the designated recipient with higher probability of transmission of positive messages. This effect was then operationalized as the MUM Effect. One explanation the authors give is the negative psychological costs for the potential communicator when the message is not positive, like negative feelings (guilt, emotional contagion), concern about the recipient and societal norms (Tesser & Rosen, 1975).

Other investigations have revealed that 30% of variance in the reluctance to break bad news of psychologists can be traced back to feelings of anxiety (Merker et al., 2010). Sarason, Peterson, and Newman (1968) found that highly anxious subjects generally imitated more than low-anxious subjects in a serial recall task. Bauer and colleagues (1983) confirm this effect with their study on observation of models completing a maze by females that differed in respect to state and trait anxiety. Female undergraduates with high levels of trait-anxiety imitated more than those with low trait-anxiety levels when exposed to a model. This effect is limited to those models that are high in prestige in contrast to those low in prestige (see 2.2.2 for effects of models' characteristics in observational learning).

There is a big body of research on the contagious effects of emotions expressed by a model during observation on the observer's perceived emotional state. For a more elaborated description on the phenomenon of *emotional contagion*, see 2.2.2.1.

But not every person experiences the same level of stress before, during and after difficult situations. This relation has been investigated for physicians BBN which is an activity connected to negative emotions (Ptacek et al., 2001). According to Parle and colleagues (1997) there are three cumulative factors that influence the communication of physicians with cancer patients. On the one hand the conversations are affected by *knowledge and skills* acquired by the doctor beforehand, like interviewing steps. But there are more aspects

that are crucial, like *attitude and beliefs*. While the belief that someone close could die has shown to have an inhibiting effect, the authors do not recommend working therapeutically with the doctors concerned. Rather, it is sufficient to identify and challenge the beliefs. The expectancy of an uncontrollable situation with extremely emotional patients and their own emotional involvement is a belief that often prevents health professionals from further engaging in BBN conversations. It has been shown in previous studies that physicians' anxiety levels correlate negatively with their clinical skills (Turner, Helper, & Kriska, 1974). In situations that are especially wearing for physicians, like BBN, it is therefore essential to acquire communicative and emotional skills (Anolli et al., 2006).
The third influencing factor is the *support* received from the workplace. Practical and psychological support has a significant effect on the application of BBN skills.

It can thus be concluded that the acquisition as well as the performance of complex skills are closely linked to the emotions experiences by the learner. Different levels of anxiety during the learning process as well as outcome should therefore be assessed in the scope of interventions targeting skill acquisition. The assessment must consequently include instruments that go beyond knowledge-based testing.

2.1.3 Measurement of Skills

A method to assess the change in skills is simulation. Simulations can be screen- or video-based, include Standardised Patients, or role-play (Lane et al., 2001) and are applied extensively in medical education. At the same time simulations can reduce negative expectations and give the opportunity to practice complex skills like difficult communication in a safe environment.
Another prominent method to assess complex skill in medical education are Objective Structured Clinical Examination (OSCE) stations, first introduced by Harden, Stevenson, Wilson, Downie, and Wilson (1975). While OSCEs are widely accepted as an instrument to test complex clinical skills and reliability has been confirmed in reviews (Patrício et al., 2013), correlation of scores obtained by the OSCE performance and written examinations are often only weak (Fischbeck et al., 2011). On the downside there are high costs of time and money for all simulation-based assessment of skills.
But even high-fidelity simulations will still be somehow different from the situation simulated (Hanna & Fins, 2006; Issenberg, McGaghie, Petrusa, Gordon, & Scalese, 2005) which could lead to reduced levels of anxiety during performance and have a negative effect on transfer. The latter concern seems to be negligible as reviews on effects of simulation-based medical education on patient outcomes (Zendejas, Brydges, Wang, & Cook, 2013) and medical students knowledge and skill acquisition (Laschinger et al., 2008) indicate small to medium positive effects.
Concerning the affective state during simulations, there are studies that hint towards simulation as leaving learners in an insecure affective state which could hinder transfer to later performance in real world settings. Students working with a rather realistic simulation (SlimMan, a medium fidelity whole-body manikin simulator that was used to simulate cardiac and respiratory emergencies) described feelings of anxiety (Paskins & Peile, 2010). In

another study, nurse practitioners' feeling of confidence was only enhanced by learning to manage a cardiac event with case studies and not when learning with a simulation mannequin (Y. K. Scherer, Bruce, & Runkawatt, 2007). Still, besides the rather limited power of those two studies (the number of participants was very small: 23 and 28 participants, respectively), it is noteworthy that realistic simulations constitute an instructional approach that may evoke the same (negative) emotions like the real world setting simulated.

In summary, it can be said that there are positive effects of simulation-based motor-skill trainings. The next step should be the broadening to other domains like communication.
Generally, simulation should be supported with feedback and guided practice to ensure acquisition of the targeted skills. A prerequisite is the clear definition of the often highly complex skill that is to be learned (Salas, Rosen, Held, & Weissmuller, 2009). It is unclear whether simulation should be applied before meeting real patients or in a later stage of learning a new skill (Issenberg et al., 2005). To construct learning environments with simulation that still lead to good performance albeit inducing anxieties a general distraction of the learners by being too anxious or ashamed has to be avoided in order to fully concentrate on the task.
Following the mapping of cognitive and affective aspects of skill acquisition and the description of typical methods applied to foster complex skill acquisition in medical education, a more comprehensive overview of instructional approaches that have been shown to be effective in problem solving will be introduced with the concept of example-based learning. This term comprises different approaches and can include features of simulation.

2.2 Example-Based Learning

During the second phase of VanLehn's model of skill acquisition (1996), following the phase of knowledge integration in existing structures (see 2.1.1), problem solving can be enhanced with examples by guiding the learners' attention to relevant aspects and develop appropriate schemas (Crippen & Earl, 2007; Schworm & Renkl, 2007). There are more ways to design instruction, but because of the popularity example-based learning has among learners, worked examples are one of the most applied and at the same time most effective instructional approaches in terms of more learning in less time (Salden, Koedinger, Renkl, Aleven, & McLaren, 2010; VanLehn, 1996). An article by Renkl (2014) presented a new theoretical framework for learning with examples that is very much oriented towards instruction. Under the concept of example-based learning not only worked examples but also learning with modelling examples is included[7]. Both approaches have an overlap in some relevant areas but also specific differences. The two instructional approaches have also been compared and discussed by van Gog and Rummel (2010) who found - besides the undeniable differences - a big body of consent.

[7] The theory of example-based learning by Renkl (2014) also includes findings from a third research area, analogical reasoning. This additional aspect will not be taken into account in this work.

Like worked examples, observing a modeling example has been shown to foster learning in academic contexts. The concept of modelling was introduced by Bandura as observational learning and is predominant in social learning. The main new aspect of Bandura's theory, compared to the then dominant behaviourism, is that learning by doing is not the sole mechanism to acquire new behaviours but observing others can lead to learning as well. While vicarious learning replaces costly and time consuming trial and error strategies (Bandura, 1986), learning through modelling may not be confused with mere imitation. Observed behaviour is not exactly copied but rather integrated in existing prior knowledge structures and behavioural patterns.

In general, a real performance includes many more cues than a theoretical schema, classical worked example, framework, or protocol can convey. When special performances like motor skills or communication are the learning target, observational learning has typically been applied. Contrary to early work in this field, this can also be implemented by observing worked examples. To further elaborate on the mechanisms of modelling and the advantages that can be brought to worked examples by including modeling examples relevant literature and empirical work will be reviewed in chapter 2.2.2 to shed light on the aspects of the case that is presented as example. The relevance of the modeling example case is underpinned with findings from research targeting empathic reactions towards others (2.2.2.1), operationalised in media research as parasocial interaction (2.2.2.2).

The selection of the case is one out of four aspects that have been identified to be crucial for learning with examples (Renkl, 2014). In order to conclude on the applicability of example-based learning for skill acquisition, all aspects will be studied more closely. Generally, research has attested worked examples positive effects in connection with a reduction of cognitive load that uncovers higher volumes of cognitive capacity to apply on the targeted problem solving steps. This effect has been labelled the worked-example effect (Sweller, Van Merriënboer, & Paas, 1998) and will be elaborated further in chapter 2.2.1.1 to conclude on the state of the art as well as come to practical implications in respect to the selection of assessment thereof.

Other characteristics are the number of examples provided for the learners and the connection of those examples to the principle behind the solution steps that can be explained by the learners to themselves. Therefore, chapter 2.2.1.2 will discuss the emergence of self-explanations to conclude on most effective instructions. Besides perfect examples, there are also positive effects of learning with erroneous examples. Chapter 2.2.1.3 will describe possibilities and limitations to learn complex skills by working with examples that include incorrect solution steps.

The need to support the learners in order to connect the principles with the examples displayed is another relevant aspect and will be discussed in more depth to conclude on possible amelioration of example-based learning (2.2.3).

2.2.1 Worked Examples

Worked examples have been investigated by many researchers and mostly been found to be effective in enhancing learning compared to other instructional approaches. The construction of schemata is supported and becomes more and more automated. This positive effect holds true under certain conditions such as a relatively low level of prior knowledge (Atkinson et al., 2000). In addition, worked examples have been found to be effective only if the learners are actively involved. This can be fostered by the addition of a practical task (Cooper & Sweller, 1987; Sweller & Cooper, 1985) or the request to invest in self-explanations during the preoccupation with the example (Crippen & Earl, 2007), see 2.2.1.2. One mechanism of example-based learning is the avoidance of overload in the working memory with inappropriate problem solving strategies (2.2.1.1). Trial and error strategies are avoided by pushing the learners into the right direction (Renkl, 2014).

Besides differences in content and design - including different types of media used - the basic structure of classical worked examples consist of three parts: 1. Problem statement to foster the awareness of the problem/task to be solved, 2. Procedure how to solve the problem in several steps that follow a given order, and 3. The solution of the problem. Those elements can be presented by schematic representation, a mathematical formula, as well as by a model to be observed (see 2.2.2). After studying the worked example, the learners are usually asked to solve the same or a similar problem themselves (Atkinson et al., 2000) to ensure transfer.

But while worked examples are often combined with problem-solving to make the learners apply the learning content immediately, a recent study found that the effect of worked examples is not significantly enhanced by the addition of problem solving tasks. Van Gog, Kester, and Paas (2011) compared four different instructional conditions with two factors examples only vs. example-problem pair and problem solving only vs. problem-example pair randomly assigned to 103 secondary education students from schools in the Netherlands. They were all novices in the task of applying Ohm's law to electrical circuits. During the experiment, the students had to work with two paper-based malfunctioning parallel electric circuits while being provided with the formula of Ohm's law. The problem format consisted of three questions / requests to be answered by the students. The example format consisted of the answers to the three questions to be studied. The post-test consisted of two problems of which one was very similar to the training material, while the second problem included two faults instead of one. Mental effort, indicating cognitive load, was measured with the mental effort rating scale by Paas (1992). During the 30 minutes session, the participants first completed the prior knowledge test, and then worked with the training cases including a mental effort scale, respectively. Time on training task was limited and the order of the tasks could neither be changed, nor were the students allowed to look back to already finished tasks. After handing in the test task sheets, the students were provided with two test tasks plus mental effort scale.
Results showed that the invested mental effort was significantly lower in the example only condition as well as example-problem conditions. The same pattern was found for performance on test tasks with higher ratings for example only and example-problem conditions.

This study shows that learning with problem-only and problem-example-pairs is less effective in respect to performance in a post-test and leads to more mental effort than learning with examples. In sum, learning with problems, compared to examples is less effective and examples are only effective when presented before the problem.

The main domains where worked examples have been implemented and studied so far are math, physics and computer programming. Problem cases from those fields are typically well structured, distinctly defined, and investigated in laboratory studies. More recent research also applies worked examples in less structured fields such as the recognition of designer styles (Rourke & Sweller, 2009) and argumentation skills (Schworm & Renkl, 2007).

There are principles for the design of worked examples to make use of their full potential. Eight design guidelines have been identified in a review of empirical studies by Shen and Tsai (2009) and should help to enhance the effectiveness of learning with worked examples (see Table 1). Those principles are closely related to the concept of cognitive load (2.2.1.1).

To be effective, the worked example has thus to focus on the basic principle that is needed for the solution of the problem type in question and ensure transfer to later performances. To support the application of acquired knowledge, examples can be made incomplete so that the learners are forced to invest in problem-solving activities already during studying the worked example (Renkl & Atkinson, 2003). In later performances it is important to further advice the learners by giving feedback on the application of acquired knowledge. Especially novice learners need feedback when learning with worked examples. The reason is the inability of novices – especially those scoring in the lowest quartile - to self-assess their performance accurately (Dunning, Johnson, Erlinger, & Kruger, 2003; Pirolli & Recker, 1994; van Gog et al., 2011). Hodges, Regehr, & Martin (2001) confirmed this effect in the domain of medical education by investigating the self-assessing skills of 24 medical students having a BBN conversation with a Standardised Patient. Video-taped conversation of students BBN were rated by them as well as by experts rated before the students were shown four other BBN interviews that differed in quality. After watching, they had the chance to re-score their own performance. In contrast to the highest scoring third (expert rating) that underestimated their performance at first and were then able to calibrate their estimation quite accurately, the lowest scoring third (expert rating) still highly over-estimated their performance.

Besides the competence level of the learners, the complexity of the instructional design is another aspect that makes feedback especially crucial. Therefore, learning with examples that are more demanding, like erroneous examples, needs to be supported with feedback (see 2.2.1.3). The abstract rules and principles underlying the worked example need to be identified by the learner. This learning process is facilitated by examples that are adapted to the learner's level of prior knowledge and can be augmented by providing realistic examples which lead to more transfer.

Table 1: Eight design principles for worked examples taken from Shen and Tsai (2009).

Principles	Explanation
1. imagination principle	Depending on prior knowledge it is helpful to introduce procedures and concepts (Ginns, Chandler, & Sweller, 2003).
2. completion principle	Incomplete worked examples support the quality of self-explanations and close and middle transfer, depending on structure and surface (Renkl, Atkinson, Maier, & Staley, 2002).
3. fading principle	The successive integration of problem solving by stepwise deleting the respective last solving step of the worked example connects example study and problem solving. Combined with self-explanation close, middle and even far transfer can be supported (Renkl et al., 2002). Adaptive fading is superior to static fading and the interaction of prior knowledge and pace of transitioning needs to be considered (Reisslein, Sullivan, & Reisslein, 2007).
4. process principle	Subgoals are superior to steps that a) highlight important aspects b) label the solution c) introduce basic structure of the target, and d) lead learners to discover meaningful generalisations (Catrambone, 1994). In addition, molar (smaller) subgoals are superior to modular subgoals (Gerjets, Scheiter, & Catrambone, 2006). Generally, worked examples should concentrate on *why* and *how* (van Gog, Paas, & Van Merriënboer, 2004).
5. presentation principle	Integrated material enhances learning and prevents split-attention effects (Sweller et al., 1998).
6. media principle	Integration of visual and verbal is superior but only in simple problem cases. A human voice is superior to a synthesized voice (Mayer & Moreno, 2003).
7. timing principle	Integrated instructions with a combination of problem and worked example are superior to problem only (Stark, Gruber, Renkl, & Mandl, 2000).
8. self-explanation principle	The overwhelming amount of information of worked examples can be better handled in the combination of worked examples and (tailored) self-explanation prompts. This leads to enhanced performance, problem solving, skill, self-efficacy and motivation (Crippen & Earl, 2007).

It can thus be concluded that worked examples have shown to be an adequate and effective tool for the acquisition of skills for learners with low prior knowledge under some conditions. Compared to the level of performance the skill to be learned can be categorised as complex.

2.2.1.1 Cognitive Load

The cognitive architecture is limited through working memory capacity. This needs to be taken into account when selecting instructional approaches and designing learning material. Especially novice learners with little prior knowledge in the relevant domain easily suffer from too much information and activity at a time. By providing those learners with instructional support superfluous cognitive load can be reduced and working memory capacity can be focused on relevant procedures.

Originally, cognitive load theory consisted of two sources: Extraneous load being associated with aspects that are not necessary to solve a problem and occupy cognitive resources that otherwise could be allocated to task-relevant aspects. Intrinsic load on the contrary, describes the activity of the working memory linked to the actual task. The more relevant prior knowledge a learner has, the less Intrinsic load is necessary for problem solving (Sweller, 1994). Both, intrinsic and extraneous load may not exceed a certain level to enhance learning. On the contrary extraneous load has to be minimised and intrinsic load should be optimised by selecting optimal learning material for the learner's prior knowledge. After conceptualising the cognitive load theory out of those two components, a third type of load has been added later on (Sweller et al., 1998), which is why it is now also labelled the triarchic theory. Germane load has been introduced to measure the 'intentional cognitive effort' (Kalyuga, 2011) during problem solving that is directly linked to the working memory resources that deal with intrinsic load (Leppink, Paas, van Gog, van der Vleuten, & van Merriënboer, 2014).
Besides already obvious interrelations between the three aspects of cognitive load, Sweller (2010) recently proposed that element interactivity, the mechanism that was proposed to underlie intrinsic load (Sweller, 1994), also explains extraneous and germane load. Element interactivity differs, depending on the level the aspects to be learned (=elements) refer to each other.

The worked-example effect is explained through different processes on the cognitive level that occur while learning with worked examples in contrast to learning by problem solving. By providing the learners with solution steps, means-end search is prevented and the acquisition of cognitive schemata is supported (Sweller, 1988; Sweller et al., 1998). Learning with examples is hence effective because extraneous load is reduced (Schwonke, Renkl, Salden, & Aleven, 2011).
Besides the learners' characteristics like prior knowledge, the complexity of the learning material, and their interaction (Paas & Van Merriënboer, 1994), a recent framework additionally includes the physical learning environment as causal factor for cognitive load (Choi, van Merriënboer, & Paas, 2014). Choi and colleagues (2014) offer practical implications by arguing that the physical learning environment, like fresh air in the learning room or loud noise during an intervention moderate the effects of instructional approaches. Moderation can take place on three levels, by influencing the learners a) cognitions (e.g. shift of attention through irrelevant acoustic inputs), b) physiology (e.g. lower oxygen

saturation through hot temperatures), and c) affective state (e.g. anxiety provoked by highly evaluative situations).

The main declaration of cognitive load theory is the adverse effect of cognitive load on learning. One finding is the negative effect of highly complex learning material and/or too much control for little experienced learners on learning. This effect is attributed to enhanced levels of cognitive load. In a study that varied the source of illustrations (learner generated vs. instructor generated) learning with illustrations provided by the instructors had positive effects on understanding, reduction of cognitive load, and perceived task difficulty (Schwamborn, Thillmann, Opfermann, & Leutner, 2011).
Interestingly, this effect was not found in the studies by Homer, Plass, and Blake (2008). In two studies, learning with video, operationalized as multimedia learning environment with video of a lecture, added to the associated slides, was compared to learning with slides only combined with the lecture's audio track. Those who learned with the videotaped lecture experienced more cognitive load, measured with the questionnaire by Paas, Van Merriënboer, and Adam (1994). But higher cognitive load did not result in less learning operationalized as recall and transfer. The missing correlation could not be explained by social presence (degree of connectedness to the speaker). More personal learner characteristics were analysed during a follow-up study with the same experimental variation and dependent measurements, complemented with a questionnaire to assess visualiser/verbaliser learning preferences. There were no main effects of learning preferences but significant interactions with those who prefer visual learning reporting more cognitive load when learning without video and those who prefer verbal learning reporting more cognitive load in the video condition.

The general problem is how to measure cognitive load. By each measuring instrument, a slightly different definition of the concept is attended. Generally, performance-based measurements, subjective ratings and psychophysiological indices have to be distinguished. Also, there are on the hand instruments to measure overall experienced cognitive load, while on the other hand more specific instruments aim at measuring the different types of load. The drawback of the instruments distinguishing between the three types of load is that they are not clearly definable. Load experienced by a novice learner as extraneous might well be categorised as germane load for an experienced learner (Kalyuga et al., 2003). This dilemma is known as the expertise reversal effect. A typical subjective instrument is the rating scale of perceived task difficulty tested by Paas and colleagues (1994). The invested mental effort had to be estimated on a scale ranging from 1-very, very easy (low mental effort) to 9-very, very difficult (high mental effort) and was compared to the physiological analysis of heart-rate variability in respect to reliability and sensitivity. The authors report that of both instruments that are constructed to measure mental effort, the rating scale yields better results in respect to reliability and sensitivity.
DeLeeuw and Mayer (2008) intercorrelated three different instruments and also analysed relationships to three outcome measures response time, intrinsic processing, and germane processing. While the correlation of the instruments, namely a self-report scale to measure

mental effort, a secondary visual monitoring task to measure response time, and a difficulty rating was rather low, each of the instruments correlated with one of the outcome measurements, supporting the triarchic theory of cognitive load. Precisely, germane processing was correlated with difficulty ratings while intrinsic processing correlated with effort ratings. Extraneous processing was related to the response time measurement.
Other recent studies involve neurophysiological measurements such as electroencephalography - EEG to complement subjective data from questionnaire-based instruments (Antonenko, Paas, Grabner, & van Gog, 2010). But so far, guidelines to interpret and apply those new possibilities are still missing. Another approach that appears to be less intrusive is the measurement and analyses of linguistic patterns that can be run as a real-time assessment (Khawaja, Chen, & Marcus, 2014).

This controversy of definition and operationalization of cognitive load has been criticised by many researchers. Still, it is a central concept in the instructional approach of example-based learning and has to be included for the sake of completeness. It can generally be concluded that research on cognitive load, that is in most cases very well controlled, supports the efficiency of worked examples for novice learners. The basic mechanism is that learner characteristics like prior knowledge and learning preferences interact with the learning material. This relationship should be analysed further by broadening the focus to more aspects of the learners. Furthermore, moderating effects of affective aspects should be included.
Until now there is no empirically supported solution of how to measure cognitive load and its scales, only hints by a comparative study that favours rating scales over heart-rate measurement.

Besides fitting levels of cognitive load, another crucial condition for effective learning with worked examples is the activity of self-explaining during learning.

2.2.1.2 Self-Explanation Effect

Actively explaining a worked example to oneself is critical in the scope of learning with examples and originated the label *self-explanation effect* (Chi et al., 1989; Hilbert & Renkl, 2009; Renkl, Stark, Gruber, & Mandl, 1998; Shen & Tsai, 2009; VanLehn, Jones, & Chi, 1992).
To explain a problem or the solution steps of an example is a learning strategy that can help the learner to develop a deeper understanding of the content. Crippen and Earl (2007) even state that worked examples alone hinder good learning results in terms of performance, problem solving, skill, and self-efficacy. Only the combination of examples with self-explanations results in enhanced learning outcomes. But what is a self-explanation and why does it foster learning?

According to Chi and colleagues (1989) self-explanations are explanations by the learner that are reasoned when solution steps in a given example are poorly elaborated. Cognitive mechanisms that are applied by the learners can be manifold and depend on the precise structure of the learning material as well as on the prerequisites of the learner. In general, self-explanations infer on the *how* and *why* of the learning material presented (Siegler, 2002) and 'go beyond the information presented in the example' (VanLehn, 1996)(p.523). There are learners that naturally explain to themselves spontaneously while others have to be supported. To correctly self-explain, the learner has to understand the mechanisms of the process presented. Experience or prior knowledge is the basis of learning with self-explanations, as Lombrozo (2006) states in her review on explanations: When learning by self-explaining new information is integrated into existing beliefs.

General strategies while self-explaining are to infer surface features to the problem's structure and integrate the provided information into already existing prior knowledge. If the learner's prior knowledge is faulty, the revision can be self-explained (Roy & Chi, 2005). More explicit, self-explanations can consist of conclusions and predictions deducted from prior knowledge (Lenhard, 2009; Stark, 2000), or statements about conditions and consequences (Aderhold, 2008).

Chi and colleagues (1989) were the first to talk about the positive learning effect of self-explanations on learning outcomes, the so called self-explanation effect. This effect cannot only be explained by enhanced time on task (Renkl, 1997) and was first attributed to the learners' inference generation by pausing and explaining the rationale behind the examples to themselves. Pirolli and Recker (1994) made a direct link between self-explanations and metacognition, claiming that the difference between good and poor learners can be traced back to the enhanced use of metacognitive processes by the successful learners. Later, Chi (2000) revised her initial definition stating that the effect self-explanations have is due to two separate factors. On the one hand the generation of inferences and on the other hand the repairing of the learner's own mental model if a divergence between the learner's own representation and the model displayed is perceived. This divergence often appears and is intensified when learners are confronted with erroneous examples. Examples that are not completely correct cause less principle-based self-explanations and instead new kinds of explanations that can be effective as well are given (Große & Renkl, 2007). But besides those explanations, Lombrozo (2006) states in her review that the mechanism how explanations work are not yet fully uncovered. One critical aspect is the correctness of the learner's prior beliefs and knowledge the new information is to be integrated with. If the prior knowledge is not correct, the challenge changes and becomes more straining (Kuhn & Katz, 2009). It becomes more difficult to detect the causal connections between the result and the preceding processes.
In general, the type of explanations given most often is causal explanations (Lombrozo, 2006). According to Chi (2000) this is also the most effective type of self-explanations.

Renkl (1997) was able to predict learning by assessing the quality of self-explanations of university freshman students. Those students whose self-explanations included principle-based explanations, explanations of 'operator-goal combinations, and more anticipative reasoning' learned most. The self-explanation effect has also been shown to be effective in multimedia learning, for example with video-based examples presenting the complex skill of argumentation (Schworm & Renkl, 2007). Those contexts complex and self-explaining has to be supported for some learners as otherwise learning might be retarded. Roy and Chi (2005) recommend to 'provide complementary information' (p. 283) in order to support and guide the learners in making inferences and generate relevant self-explanations.

Lombrozo (2006) states that the generation and evaluation of explanations are spontaneous. Other authors are less optimistic and write that self-explanations have to be guided because they do rarely occur unprompted (Chi, de Leeuw, Chiu, & LaVancher, 1994; Renkl et al., 1998). While it is undeniable, that the ability to infer causal relations is a 'basic property of human beings' (Siegler, 2002, p. 37) and can already be found in very young children, it depends on the problem as well as on prior knowledge if learners succeed in self-explaining later in life. To support learners in their natural driven attempt, learning material can include sections of missing information to trigger the learner to make inferences to fill in the gaps and explain this process. As writing or typing self-explanations can constitute a distraction in the learning process, Roy and Chi (2005) recommend to ask the learners to verbalise their explanations instead.

So on the one hand, self-explanations are a personal characteristic that can be regarded as relatively stable (Renkl, 1997), while on the other hand the elicitation of self-explanations can be fostered and is effective in respect to transfer (Bielaczyc, Pirolli, & Brown, 1995; Renkl et al., 1998; Roy & Chi, 2005).

Pirolli and Recker (1994) give cause to consider other aspects beyond self-explanations to have an effect on learning, stating that the positive effect of self-explanations does not in any case lead to enhanced performance. In their experimental study, they found significantly more self-explanation by learners doing well in knowledge tests when splitting everyone into good and poor performers. On the contrary there were only small effects on learning when the authors run more detailed analyses of individual differences.

There are as well empirical findings that suggest that self-explanations may even hinder learning. In a study by Kuhn and Katz (2009) fourth graders performed better on a transfer task on causal inferences after investigating a database compared to those who additionally had been asked to generate self-explanations. The students worked with an Earthquake Forecaster software and learned which features had what kind of effect on the outcome (here: earthquake) by manipulation. As pre-test the students worked individually with the software and constructed their own case. After the outcome of their simulation was presented they were asked to make inferences in respect to causalities. Subsequently they were assigned to pairs and worked with another case on ocean voyage that was structurally similar to the earthquake forecaster. They were then asked to make a prediction and state the influencing variables. In the explanations condition, students were additionally asked to explain the causal effects. There was a general positive training effect (pre-post) of the intervention. In addition, the students from the no explanation condition spent less time on the task and scored better in the post-test which was identical to the pre-test. The authors conclude that learners should not only be prompted to explain but that they should learn 'to evaluate whether those explanations are correct' (p. 393). Berthold and Renkl (2009) come to the same conclusion. In their study high school students who studied worked examples from probability were supported with self-explanation prompts. While there was a general positive effect of the support, some students produced incorrect self-explanations that had a negative effect on their procedural or application-oriented knowledge acquisition. Siegler (1995) avoided the danger of incorrect self-explanations at least partly by providing children with the correct answer to a question and subsequently asking ‘How do you think I knew that?’. Children from this *explain-correct-reasoning condition* resulted in more learning compared to the group that only received feedback on whether their own answer was correct (*feedback-only*) and the third group who were asked for their reasoning before receiving feedback on the correctness (*explain-own-reasoning*). The effect was larger the more difficult the questions to answer were. In a follow-up experiment with third- and fourth-graders, Siegler (2002) found that it is beneficial for learning if not only a correct solution is self-explained, but in addition the children are asked to also explain why another answer is not correct. The author explains this effect by the weakening of underlying incorrect strategies in addition to the strengthening of underlying correct strategies.

In the context of modelling an aspect similar to self-explanations is described, labelled ‘coding’. This label describes the organising process during the observation of a model when observers ‘reduce the diverse elements […] into a pattern of verbal symbols’ (Decker, 1980, p. 628). Learners - observing a model - who spontaneously either code verbally or numerically can reproduce the modeling behaviour better and show more transfer of the behaviour than those who did not code (Decker, 1980). Coding can be formalised by the introduction of learning points to the observing learner. Those learning points contain a description of the behaviour (behavioural), the essential elements of the behaviour (summary label), or the principles behind the behaviour (rule-oriented) and can enhance reproduction and also generalisation of the modeling behaviour compared to modelling without learning points (Decker, 1980, 1984). Rule-oriented learning points are the most effective in respect to generalisation and transfer (Decker, 1984). Still, this is not exactly the same concept and while there are some studies that investigate self-explanations in modelling situations, Renkl (2014) noted in his overview paper that the cases analysed are not typical for observational learning and thus cannot be generalised.

Concluding, it can be said that research has shown that learners can be supported in self-explaining with causal explanations being most beneficial. Still, there is not always a direct positive effect on learning as explanations may be incorrect. In the field of modelling the role of self-explanations is not clear yet and needs to be investigated further. Especially Siegler’s (2002) finding that the explanation of erroneous solutions is beneficial for learning should be investigated in the context of observing coping models[8]. Generally, learners can be supported in their attempt to explain an example to themselves with self-explanation prompts (Roy & Chi, 2005).

A special subcategory of examples that might provoke higher values of cognitive load and must thus be well supported is erroneous examples that present incorrect solution steps, a coping model, or wrong final solutions. Those examples have though been shown to be

[8] Coping models, opposed to mastery models, do not perform totally correct but make some mistakes and sometimes include coping mechanisms to deal with erroneous behaviour (see 2.2.1.3).

effective under some conditions and sometimes even foster more learning than correct examples.

2.2.1.3 Erroneous Examples

The acquisition of complex skills by example-based learning has been mostly investigated with proper cases, like the video-based examples to foster argumentation skills in teacher students by Schworm and Renkl (2007). But besides presenting learners a perfect solution to a problem, researchers have also investigated in how far learning from erroneous examples can be effective. The idea is based on VanLehn's (1999) assumption that learners confronted with a dead end of an approach reflect more on the solution and is in line with the search for a perfect model as described above in the context of observational learning (see 2.2.2). Enhanced preoccupation can ultimately lead to better learning. This positive effect has been verified by different studies (Große & Renkl, 2007; Stark et al., 2011) but has been found to be true only under some preconditions: Solely learners with high prior knowledge were able to benefit from erroneous examples in the study by Große and Renkl (2007).

In a study in the context of medical education, Stark and colleagues (2011) focused on the effects of feedback on learning with erroneous examples. They compared the effects of case-based examples (erroneous vs. without errors) and feedback (elaborated vs. KOR – knowledge of results) on diagnostic competence, cognitive load, and time on task. 153 medical students were randomly assigned to one of the 2x2 design conditions. In the pre-test domain-specific conceptual knowledge was tested with a multiple choice test that was also applied as part of the concept *diagnostic competence* in the post-test. This measurement was complemented with ten key feature problems to measure strategic knowledge and three problem-solving tasks to measure conditional knowledge. Additional measurements were nine items for cognitive load and time-on-task.

In the training phase, each student saw six worked examples of student working out a diagnosis and receiving feedback from an expert. The erroneous example included up to five errors and a wrong diagnosis by the protagonist while no errors were included in the *without errors* condition. Still, all students learn the correct procedure as every error is corrected by the expert's feedback subsequently. The elaborateness of this feedback is varied in the feedback factor. Students in the KOR condition received feedback on the correctness of the diagnosis and procedure while the elaborated feedback included explanations and references to underlying knowledge.

Results showed that students who received elaborated feedback spent significantly more time-on-task but there were no effects of time-on-task on diagnostic competence. More time-on-task was most beneficial for those who worked with correct examples and received KOR feedback. Prior knowledge had a large main effect on diagnostic competence with better results for those with higher prior knowledge. There was an interaction effect of the two factors on diagnostic competence as well as on strategic and conditional knowledge but not on domain-specific conceptual knowledge. Post hoc test revealed significantly better results for the combination of erroneous examples with elaborated feedback compared to KOR feedback. No effects of the feedback factor steps in the condition without errors. There was no main effect on cognitive load, only an interaction effect with higher values for those who worked with erroneous examples and received KOR feedback. Prior knowledge was negatively correlated with cognitive load which in turn was negatively correlated with diagnostic competence. When the more complex learning material was applied in a follow-up study, the effect of elaborated feedback was even larger, underlining the relation of feedback and complexity of learning material.

Another form of feedback can be seen in highlighting the errors in erroneous examples. Große and Renkl (2007) found a positive effect of making the errors salient on near transfer for learners with low prior knowledge. In general, intensive preoccupation with incorrect solutions can enhance learning by contrasting similar aspects and explaining the errors shown (Siegler, 2002). In an experimental study, students had the largest increase between pre- and post-test (algebraic word test) after self-explaining both correct and incorrect algebraic equations compared to those who only explained why algebraic equations were correct (Curry, 2004). Those results are supported by another study on conceptual understanding of algebra that showed best results in respect to encoding after studying a combination of correct and incorrect examples compared to the correct only condition (Booth, Lange, Koedinger, & Newton, 2013). However, there were no effects on procedural skill.

The importance of feedback on the erroneous aspects of examples is supported by research on modelling. In a study by Blandin and Proteau (2000) an erroneous example (coping model) was as effective as a perfect mastery model example when learning motor skills. Students were learning a movement pattern either by physical practice or observation of a model that was a beginner or a mastery model. The control group did not receive any kind of practice. There were no differences in respect to performance, error detection and error correction between the experimental groups. But the authors state that it is crucial for the detection and correction of errors to be informed about the level of success of the model. Otherwise erroneous reference of the correct procedure might be developed.

Bandura (1977) showed in his classic studies that the level of success a model had with his/her behaviour performed, has an effect on the observer's subsequent performance. Vicarious reinforcement resulted in higher probability of behaviour that had been rewarded in contrast to behaviour the model had been punished for which the learners tried to avoid. For more information on erroneous modeling examples displaying coping models, see 2.2.2.

It can thus be said that erroneous examples are effective under the condition of sufficient prior knowledge and feedback on the correctness of the displayed solution steps. Furthermore a combination of correct and incorrect examples can be beneficial for knowledge acquisition. Albeit the effects of feedback and prior knowledge have been studied, the relation of self-explanations during learning with erroneous examples should be analysed next, as mentioned by Stark and colleagues (2011).

An example-based learning approach that has concentrated on different learning mechanisms that may enlighten the learning process and explain further variance is modelling.

2.2.2 Modeling

'Some complex skills can be mastered only through the aid of modelling' (p. 20), like cognitive skills, speech, social interaction, but also values and attitudes (Bandura, 1986).

Bandura (1986) proposed four different processes required for learning the behaviour modeling and observed: *Attention* paid to relevant behaviour, *Retention* of the activity, the ability of *Reproduction*, and *Motivation* to reproduce the skill. Besides learning new skills, it is also possible to strengthen or weaken already existing skills. Bandura (1986) labelled those mechanisms as inhibitory and disinhibitory effects. If an observed behaviour is performed depends on the perceived ability to perform the behaviour, on the rewarding vs. punishing consequences the model experienced for its performance, and the perceived probability to be confronted with the same consequences when showing the behaviour.

The differentiation of worked examples and observational learning can partly be traced back to the approaches the respective research is rooted in. Purely cognitive research on the one hand focuses on worked examples while social-cognitive research on the other hand investigates the effects of modelling and observational learning (van Gog & Rummel, 2010). As stated earlier, this differentiation can at least partly be overcome by integrating or combine modelling and worked examples which can have positive effects on learning (Rummel & Spada, 2005; Rummel et al., 2009). Some researchers describe worked examples as expert models (Atkinson et al., 2000; Schunk, 1996) while another approach is the introduction of worked examples as content of observational learning (Rummel & Spada, 2005; Rummel et al., 2009). Both instructional approaches are most effective when applied in early phases of skill acquisition. When embedded in worked examples, the model comprised not only the person demonstrating the targeted skill but the example as a whole. Furthermore, Renkl (2014) introduces the term of *abstract modelling* labelling the distinctive features of cognitive skills modeling in contrast to motor skills. The latter can easily be imitated by the learner while cognitive skills including underlying rules and principles are not visible but must be deducted.

Observational learning is mainly applied to teach complex social skills. Problem cases in this field are typically less structured and comprise a variety of aspects what makes them more complex. Examples are complex behavioural skills, including motor skills, or negotiation (Stenning et al., 1999). Cox and colleagues (1999) compared the learning from traditional learning material like animated diagrams, texts, and worksheets and the same learning material enriched with the observation (*here: 'listening in'*) of taped dialogues of students collaboratively learning with the material. They found positive effects of the vicarious learning condition with added dialogues on understanding. Furthermore, students modelled the behaviour provided via the '*listening in*' in their discussions following the learning phase

(Stenning et al., 1999). The performance of the observed skill following an observational phase has been found to have the best effects when motor skills are modelled, as well (Blandin, Lhuisset, & Proteau, 1999).

Rummel and Spada (2005) investigated the effect of a worked collaboration examples in an experimental study and found positive effects on the application of the observed behaviour. Two instructional approaches were compared: Collaboration examples and collaboration scripts. Seventy-two advanced students of psychology and medical science (50:50) were paired to interdisciplinary dyads to collaborate in a computer-mediated setting. The dyads first completed a learning phase to acquire relevant knowledge about good communication. This was followed by the application phase during which the dyads were asked to develop a joint solution in a computer-mediated problem-solving process in the context of psychiatric cases (diagnosis, case description, and therapy plan). For the problem solving, psychological as well as medical aspects had to be taken into account.

Respectively nine dyads were allocated to one of the three conditions during which they had the opportunity to collaborate on solving a case (case 1) and the control condition not taking part in the learning phase: a) observation of a worked collaboration example (model condition) with self-explanation prompts, b) scripted collaboration (script condition): paper-based, with instructions to foster the same elaborative behaviour presented in the model condition, c) unscripted collaboration (unscripted condition): free collaboration of the dyads without further instructional support, and d) no learning opportunity (control group). Conditions a) and b) both included all relevant aspects of the targeted collaboration that consist of coordinative skills, communicative skills, and domain-specific skills. In addition, psychology students were provided with information on psychological aspects while medical students received information on medical aspects of diagnosis as well as therapy.

During the problem-solving the dyads were in separate rooms and worked on a new case (case 2). All dyads communicated via a video-conferencing tool including audio, shared text chat, and a live video-picture of the partner but without any further instructional support. Following the implementation phase, those three skill-levels were assessed during the problem-solving process on the basis of log files (detection of individual vs. joint work with focus on individual phases), dialogs (analyses in respect to coordination and communication), the quality of the joint solution (comparison to an expert's diagnosis and therapy plan on a quantitative level), and a post-test (subscales on metacognition relating to *good communication* and the construction of a *good therapy plan*).

The results of the log files revealed that deviation of individual phases (duration and distinctiveness) during the problem solving from the optimal process was significantly smaller in the model and the script condition compared to no script and control condition. The dyads in the model condition spent the largest amount of individual work and thereby differed least from the exemplary collaboration. There is a significant main effect of the conditions on the quality of joint solution with better diagnosis in the model condition contrasted to the script condition and better therapy plans in the experimental conditions compared to the unscripted and control condition. The instructional conditions also outperformed conditions c) and d) in the post-test. The dyads with good joint solutions were

mostly allocated in the group that spent high amounts of individual work which underlines the importance of individual work phases. As a side-note the authors report that the dyads in the model condition collaborated with more enthusiasm than did the ones in the script condition.
Overall, this study confirms the hypotheses that a worked collaboration example can foster good communication skills and is at least as good as learning with a script. The non-scripted condition did not differ substantially from the control condition.

In a subsequent study Rummel and colleagues (2009) compared the efficiency of modelling examples with learning with scripts with the same rationale but with a new rating scheme that leads to deeper matter analyses. The rating scheme includes nine dimensions and aims at indicating on the one hand the complexity of the communication (qualitative aspect) and on the other hand allocates ratings on the basis of the qualitative impression (quantitative aspect). Additionally, the instructional approaches were combined with elaborational support including prompts and reflective self-explanations which resulted in a 2x2 design and a control condition. The authors report positive effects of worked collaboration examples that were increased by the addition of elaborational support. Interestingly, participants from the model conditions also rated the learning phase to be more helpful than did participants from the script conditions (Rummel et al., 2009).

Both studies underline that complex social skills (here: collaboration of psychology students and medical students) can effectively be learned by modelling, even with just one exemplary case. Learning with a worked collaboration example has been shown to be more effective in this context than learning with collaboration scripts.

There are opposing views and findings in literature on what kind of model is the most efficient for learning. Bandura (1986) states that those models are more influential that are successful, appear to be similar to the learner, and are perceived as likable. Other critical aspects for observational learning are the distinctiveness of display, functional value of behaviour and if the model was reinforced (Bandura, 1977).
According to Schunk (1996) the role model should be of the same ability level as the observer to reach highest levels of self-efficacy. On the other hand Baum and Gray (1992) found supervisory models to have a more positive effect compared to a self-observation or a novice on video. This finding is supported by a study that compared two different models and found a significant effect on a transfer test: students who observed an expert example performed significantly better than those who observed an advanced student model regardless of their own expertise (first-year vs. second-year physiotherapy) (Boekhout, Van Gog, Van De Wiel, Gerards-Last, & Geraets, 2010).
Bauer and colleagues (1983) found that models high in prestige were imitated more than models low in prestige when completing a maze. The high prestigious model was introduced as an expert and technical advisor with doctorate in experimental psychology while the other model was introduced as a friend of the experimenter's little sister. The effect was independent of the participants' state and trait anxiety levels. On the contrary, a study by

Blandin and colleagues (1999) did not reveal any differences between a beginner model and an advanced model on learning a new motor skill through observations.

Van Ophuysen and Hannover (2005) investigated the conditions of change in the own self-concept through observational learning. Higher levels of influence were documented of those models that appear to be likable and convey personal warmth to those who have already developed a self-concept in the relevant domain. In their study the authors compared the influence of a model playing a computer game on participants' self-concept that had either played the game before or not. In general those without prior experience were influenced more. The experienced participants were only influenced if the model appeared sympathetic to them. It can be concluded that observational learning also depends highly on the learners' prerequisites like prior knowledge. In order to ensure learning by observing a model, Bandura (1986) opted for displaying more than one model to increase probability of identification by the learner with one of the model as minimum and enhance the accuracy of reproduction (Carroll & Bandura, 1990).

Furthermore, models do – corresponding to worked examples – differ in respect to the quality of the skill displayed. Mastery models show how to behave correctly while coping models (partly) fail in their performance and often include coping strategies (Renkl, 2014).

A theory that conceptualises the relation of the learner / observer to the observed persona is parasocial interaction (PSI). It derived in media research and measures the quality and intensity of the one-sided interaction between an observer and a media persona that can either be an actor, a newscaster, or anybody alike. In the scope of their research on PSI Klimmt and colleagues (2006) found effects of the Personas'/model's gender with females provoking higher levels of PSI than male models (for more information on PSI, see 2.2.2.2).

A critical aspect for the effectiveness of modeling examples besides the characteristics of model and observer is the appraisal of the learning as highly realistic by the learner (Anolli et al., 2006; Schank, 1997). Anolli and colleagues (2006) therefore calls for 'emotional involving' learning environment that should convey a 'sense of presence' (see also 2.3).
Another critical aspect for effective learning is the resemblance of the example provided for observation and the application situation (Rummel & Spada, 2005) to ensure transfer.

So far, research on observational learning focuses on the relation of the learner (*here*: observer) and the model. In the scope of this – strictly speaking - one-sided relationship, empathy on the side of the observing party is an obvious and promising concept to take into account. Empathic behaviour can be operationalised with the concept of parasocial interaction. To elaborate if learning from modelling examples can be enhanced and moderated by empathic reactions, the concept of empathy and parasocial interaction will be described next.

2.2.2.1 Empathy

Empathy is a personality trait that is defined to be stable over time and in its tendency also across situations. Originally, empathy has been translated from the German word *Einfühlung* by Titchener (1909) and is based on the ability to identify and perceive emotions (Heberlein & Atkinson, 2009). It is a reaction that is generated when a person observes experiences of real or fictional others and identifies with them (Leibetseder, Laireiter, & Köller, 2007; Leibetseder, Laireiter, Riepler, & Köller, 2001). Empathic behaviour is sometimes equated with perspective taking (Preston & de Waal, 2002) and closely related to moral judgements and monitoring affect (Zillmann, 2006). In very young children, displayed behaviour and emotions are directly imitated and expressed without filtering. In the course of development, expression and imitation decrease and only appear under certain circumstances, like familiarity and perceived similarity with the model (Preston & de Waal, 2002).

There are two dimensions of empathic reactions that were described and investigated in parallel in the 1970s and originally seen as contradictory. One research camp defined empathic reactions to manifest exclusively on the emotional level with emotional contagion (Mehrabian & Epstein, 1972; Stotland, 1969). This approach was also supported by Bandura (1969) who referred to observers showing the same emotions like models did before and attributed this behaviour to conditioning. The other approach defined empathy as 'the imaginative transposing of oneself into the thinking, feeling and acting of another' (p. 127) and – in contrary to the first approach - excludes the sharing or contagion of those feelings (Dymond, 1949). Empathic process, in contrast to emotional contagion, takes place by taking over the perspective on an intellectual level (Dymond, 1949; Hogan, 1969). Wild, Erb and Bartels (2001) showed in their study on emotional contagion that the mere inspection of emotionally expressive faces evokes emotions in the viewers. The stronger the observed emotions on the faces shown on the pictures, the stronger were the observers' (questionnaire-based) ratings of their own emotions. There was no effect of gender and the effect was even measured when the emotional faces were only presented for a very short time span (500ms).

Today, both levels of reaction are included in the operationalization of empathy (Leibetseder et al., 2007; Leibetseder et al., 2001). A recent integrative definition is presented by Zillmann (2006) who differentiates three components of empathy: a) Disposition: Reflexive and learned skeletal-motor reactions, b) Excitatory reactions that are controlled by stimulus and reinforcement, and c) Experiential component with three monitoring subcomponents that serve the proper experience, the correction and redirection of affective reactions, and the generation of affective reactions. The experiential component is the only one that is influenced by complex cognitive mediation and happens consciously. It is a correction mechanism that can - according to the monitoring subcomponents - either lead to the decision to continue an automatic empathic reaction, alter it, or even stop it if categorised as not appropriate (Zillmann, 2006).

Unless inhibited, empathic reactions happen mostly automatically primed on an autonomic and somatic level (Preston & de Waal, 2002). But while empathy is a trait that remains more or less stable over time, there are undeniable intra-individual differences in respect to the empathic reaction in differing situations or contexts. The variance in 'empathetic reactivity' has been investigated in an experimental study conducted by Zillmann and Cantor (1977). The authors studied children's (2nd and 3rd grade) emotional responses to six different versions of a film. All films included a protagonist that showed malevolent, neutral, or benevolent behaviour. The second factor that was varied is the emotion expressed: the protagonist showed either euphoria or dysphoria. The goal of the study was to predict the affective response of the children and it's similarity to the emotion shown in the video, depending on the type of behaviour and emotion of the protagonist.

24 girls and 24 boys from a public school in the USA were randomly assigned to the six different conditions. Every child individually saw one video that was varied in a) the behaviour of the model (malevolent, neutral, or benevolent) and b) it`s expression of either euphoria or dysphoria towards the end of the video. In every case the protagonist displayed in the video was a boy of the same age as the observing pupils. In the initial scene of the video, the boy is walking down a street. He is accompanied by some other boys and subsequently meets a dog. Later on he makes himself a sandwich and is asked by his little brother to fix a toy.

The malevolent protagonist acts aggressively towards the boys as well as the dog without any reason displayed. He does not share the sandwich with his brother and, on purpose, further damages the toy instead of fixing it. The benevolent boy is acting friendly towards both friends and dog. The little brother gets half of his sandwich and receives immediate help with the toy. In the neutral condition the boy is acting neutrally towards the other boys and does not interact at all with the dog. When the little brother asks for some sandwich he offers him bread and peanut butter to make one himself and unsuccessfully tries to fix the toy.

The euphoric protagonist receives a bicycle from his parents and cheerfully rides it. The dysphoric boy also rides a bicycle but falls off and starts crying.

During watching one of the six videos the children's facial expressions were videotaped with a hidden camera. Following the video, every child was individually interviewed by the experimenter. The questions tackled a) the appreciation of the film, b) the perception of emotions, c) the affective response to the bike incident, d) the affective disposition towards the boy and e) his behaviour, as well as f) the appreciation of the bike incident. The videotaped facial expressions of the pupils were independently and blindly to the conditions rated on a scale ranging from -100 for extremely dysphoric to +100 for extremely euphoric.

The results of the validating questions confirmed the experimental set-up and the malevolent boy was liked significantly less than the neutral and benevolent boy. The videotaped facial expressions did not show any effects of the factors. The questionnaires revealed some significant results. There was a significant interaction effect of the boy's behaviour and emotion on the observers' emotional responses as well as a main effect of the boy's emotion: watching a malevolent boy being euphoric caused dysphoria in the viewers and dislike of the ending while the same emotion of the neutral and benevolent boy caused positive emotional responses as well as liking of the ending.

The authors conclude that the children first rate the model and in a second step accept his (bad) luck only if it is 'fair'. This empirical study shows that the children only experienced empathy with a character in a movie if the model was presented as being good and popular. When the model behaved in a negative way and was disliked by the children not only was the empathic reaction missing but they even showed opposite emotions (Zillmann & Cantor, 1977).

The ability to be empathic is a crucial proposition for the formation of communication, but instead of a simple correlation there is a more complex connection of empathy and communication. More empathy does not automatically lead to better communication. Leibetseder and colleagues (2001) explain that very empathic persons might be especially wearing during conversations compared to someone not at all influenced by the perspective of the counterpart. Empathy can lead to stress if the person concerned does not have strategies to deal with it – neither cognitive nor behavioural. To benefit from empathy it is important to master constructive ways to regulate emotions and avoid stress. On the contrary a repetitive confrontation with others' emotions can also lead to a reduction of the empathic reaction by an adaption to the excitatory reaction (Zillmann, 2006). But it is also possible to raise the affective response – at least for children: Prosocial video games have been shown to influences the affective response by generally shifting behavioural tendencies towards empathy – and vice versa for aggressive behaviour (Gentile et al., 2009; Greitemeyer & Osswald, 2010; Greitemeyer, Osswald, & Brauer, 2010; Howard-Jones, 2011).

In their review on communication trainings for doctors in oncology, Libert and colleagues (2001) underline the importance of adding further outcome variables to measure the quality of communication not only with rigorous schemes (e.g. communication protocols) but include for example the patient's view.
However, empathy is more than a further criteria of communication quality: Radey and Figley (2007) state that honest and essential empathy (here: *compassion*) can reduce stress and burnout in physicians. It is neither absolutely needed nor sufficient for a good conversation between doctor and patient for the doctor to be empathic. But in any case, constructive ways to regulate emotions in difficult situations are needed (Leibetseder et al., 2001).

It can thus be concluded that empathy alone is no guaranty for good communication but should be assessed with multiple methods as it is an important personal factor in interpersonal settings. The expression of empathic behaviour however, depends on situational factors as well. To support the acquisition of complex communication skills modelling examples based on guidelines for orientation can be a helpful instrument to deal with empathic reactions in a constructive way as prosocial behaviour can be fostered with video-based examples. The relation of the learner (*here*: observer) and the model presented in a video-based example, can be operationalize with the concept of parasocial interaction.

2.2.2.2 Parasocial Interaction

Parasocial interaction (PSI) was first described and conceptualised by Horton and Wohl (1956): Interaction is para-social when someone reacts to a media-persona in a way as if the persona was someone familiar and *real*. Originally, it was conceptualised as interaction between a televisions regular, like a newscaster, who uses the mode of directly addressing the viewer on a long term basis. The media counterpart or *persona* is described as being ‘ordinarily predictable’ (p. 217) in his/her behaviour as well as in respect to the time of appearance that could be integrated in the routines of everyday life (Horton & Wohl, 1956). In contrast, current work is rather based on the assumption of para-social interaction as a user experience that may arise automatically even during a one-time confrontation with a TV performer or any other mediated persona like stage actor (Hartmann & Goldhoorn, 2011; Rubin, Perse, & Powell, 1985). The interaction can be manifested not only in behaviour but also in thoughts and feelings and vary from high involvement with a persona (high PSI) to the situation that the persona is not relevant to the viewer (low PSI) (Klimmt et al., 2006; Schramm & Hartmann, 2008; Schramm & Wirth, 2010).

The concept of PSI is related to *social presence* mentioned in connection with a study of cognitive load and video (see 2.2.1.1). Social presence, or just presence is conceptualised by Lombard and Ditton (1997) as the ‘perceptual illusion of nonmediation’ which is kind of participatory experience with characters from television (Lombard, Reich, Grabe, Bracken, & Ditton, 2000). Early studies from the field hinted towards an inclusion of social factors in computer mediated settings as social presence might improve instructional effectiveness (Gunawardena, 1995). Social presence has shown to facilitate PSI with public figure on social network sites (Lee & Jang, 2013). Bandura (1986) also referred to media as a new source of ‘televised vicarious influence’ (p. 20), broadening the field of modelled behaviour.

Klimmt and colleagues (2006) name several subcategories that can be related to the three processes (cognitive, affective and behavioural PSI). Cognitive processes include attention allocation, comprehension and reconstruction, activation of prior media experience, anticipatory observation, evaluations, and construction of relations between persona and self. Affective reactions always depend on whether the viewer ethically agrees with the persona. Affective sub processes are empathetic reactions, persona-generated own emotions, and mood contagion. Behavioural processes are motor activity, physical activity, and verbal utterance.
Interaction that is para-social should not be confused with real or normal social interaction but rather be seen as an extension to it (Giles, 2002). Hartmann (2008) even suggests to replace the term ‘interaction’ by ‘processing’ to avoid conceptual confusion.

Research has shown that high levels of PSI experienced by observers result in more intense discussions and cognitions following the watching (Klimmt et al., 2006). While the concept of PSI was originally construed as positive like a perceived friendship of the viewer to the persona, more recent research has shown that also very much disliked persons on a video screen can provoke high levels of PSI. The development of friendship is based on a certain level of identification whereas PSI is not and therefore not to be equated with friendship (Dibble & Rosaen, 2011). PSI can hence be performed without having the viewer recognise similarities between him-/herself and the persona (Giles, 2002). In the extreme case of ‘socially marginalised’ groups mediated contact has even been shown to improve the viewers’ attitudes towards the personae (Schiappa, Gregg, & Hewes, 2005, 2006). This makes it less important to develop highly sympathetic models but one that is perceived as being realistic which has been found to correlate with PSI (Giles, 2002). Videos as learning material are generally more realistic that other learning materials. Therefore, the sense of realism should generally support the involvement of learners (Anolli et al., 2006).

The intensity of PSI is also influenced by characteristics of the viewer, like empathy, openness for interaction, prior experience with the persona, and the general motivation to get into contact with media figures. Besides the viewers' traits his or her current state and the attraction to the mediated persona also influence the intensity of PSI (Klimmt et al., 2006; Schramm & Wirth, 2010). However, other factors such as demographic variables are less definite/clear in their effects on PSI (Giles, 2002).

On the side of the persona, Giles (2002) names three important factors that may influence the nature of PSI: 1. Authenticity/Realism: The *perceived realism* of the persona by the viewer has been found to be a significant predictor of PSI. However, it is also possible for viewers to develop PSI with little authentic personae like cartoon characters and other fictional characters as shown by Schramm and Hartmann (2008). 2. Representation across different media outlets: Movie stars may appear in several movies as well as advertisement which makes it more likely that the viewer engages in PSI beyond the actual consumption of media. 3. User Contexts: So far, research has focused on a single media user in front of a (television screen). It is assumed by Giles (2002) that co-viewing and discussion amongst co-viewers may change the PSI developed.

This assumption is at least partly included in the design of a study by Schramm and Wirth (2010) who tested the PSI scale (Schramm & Hartmann, 2008) in three different contexts that come with different kinds of personae. In the first context (theatre) an actor was standing right in front of the audience (non-mediated) but impersonated a fictional character. The second character was mediated via television and also fictional (soap opera) while the third character was also mediated but a nonfictional participant of a quiz show. Every context was tested in a proper study. In the first context (theatre), 236 random visitors from the same play filled in the questionnaire directly after the performance. 192 participants were asked to watch an episode of a German daily soap and complete the questionnaire online. The quiz show was watched by 61 participants who completed the PSI scale subsequently. The PSI scale was reduced to 30 to 44 items (out of 112 items) to secure motivation of the participants as additional items were added to each of the contexts respectively. Realness and suitability for an ideal role model (actor in theatre), the viewer's self-confidence, perceived own attractiveness, extraversion, openness, and sociability (soap opera), empathy readiness, extraversion and preference for quiz shows (quiz show). While the focus was on the persona's features in the theatre context, the two other settings focused on the viewer's characteristics.

The scale proved to be applicable in all settings and with reduced number of items per scale, but the results and influencing factors differed between the three contexts. For the non-mediated fictional character (theatre) the models obtrusiveness and suitability for an ideal role model both had the highest impact on cognitive and affective PSI. While only the viewers' own attributes (perceived attractiveness, sociability, and self-confidence) were influencing factors in respect to cognitive and behavioural PSI in case of the soap opera, cognitive PSI with the quiz show candidates was influenced by the candidates' (personae's) attractiveness. The authors conclude that PSI seems to require the same social skills required for social interaction.

PSI has thus shown to be a pivotal concept in both mediated and non-mediated settings. Nevertheless, the number of context investigated is still limited and more contexts with different features should be tested to further validate the concept. In the context of the communication training for medical students the concept of PSI is important because (parasocial) interaction and communication are close concepts and many attributes are even similar (Giles, 2002). PSI processes, like affective reactions to the persona, can moderate the outcome of social learning (Klimmt et al., 2006). Therefore, when implementing a learning environment with video-based modeling examples for complex skill acquisition, it makes sense to include and analyse the interaction that takes place between the viewer and the model displayed (persona) in the example before the learner performs the skill him-/herself. However, other instruments should be included in order to get a clearer picture on the personal and setting-related influencing variables. Especially the differentiation of PSI from empathy is not totally clear yet.

In general, worked examples as well as modelling are most efficient when applied in cases with not much prior knowledge at hand. Furthermore, it has been shown that learning with modeling depends in its effects on the role-model presented. Despite those empirical results of a few specific studies, there is so far no research focus on self-explanations during observational learning where a skill is modelled. Also, effects of coping model and mastery model that have the same values in respect to likability relation to learner have not been investigated experimentally.

2.2.3 Instructional Support of Example-Based Learning

To support learning with examples several methods have been shown to be effective by having positive effects on the learning outcome. Expert comments and instructional elaborations are recommended by Stark, Mandl, Gruber, and Renkl (1999) when learning with worked examples. Modelled elaborations of the example can be presented to the learner during the worked example and have the function to highlight important underlying principles or give further information that supports the understanding (Stark, 1999). Besides differences in appearance and competence level of the model, the two most dominant mechanisms to support learning with examples to date target active engagement of the learners with the learning material in order to foster deep understanding. This aim is pursued by designing examples that are either a) incomplete or interrupted, or b) include prompts that explicitly ask the learner to engage in self-explaining the example. In both cases, learning is expected to be mediated by the learner's self-explanations.

Incomplete Examples

Worked examples that only show part of the solution and omit relevant steps have been found to foster transfer. This effect is attributed to enhanced elaboration processes that include working on the example in order to solve the problem presented (Renkl et al., 2002; Stark, 1999). When this instructional approach is applied, the example shown is incomplete first, forcing the learner to anticipate the missing content. In order to avoid improper reasoning to be acquired and revise incorrect reasoning, the complete solution step is presented subsequently. Stark (1999) found positive effects of incomplete examples on identically structured but differently presented problems (near transfer) and differently structured but identically presented problems (medium transfer). Furthermore, performance was enhanced for those learning with incomplete examples compared to those who worked with complete examples.

Renkl and Atkinson (2003) compared complete examples, labelled as *modeling*, against incomplete examples that were labelled *coached problem solving*[9] while focusing also on the learners level of expertise. The authors conducted a study to examine the effects of fading solution steps. Their recommendations include adaptive fading of solution steps depending on the learners' level of expertise, starting with complete worked examples for novices and processing towards incomplete tasks to be solved.

Another study by Hausmann and VanLehn (2010) compared complete and incomplete video-based examples, differentiating between two activities of the learners to explain the effectiveness of self-explanations: self-explaining as generation of content and paraphrasing as an outcome without producing new content. In the 2x2x3 design 104 voluntary learners from second semester physics courses were randomly instructed to either study a complete or incomplete example while using either self-explanations or paraphrase the example. The third independent variable was the problem with three different problems involved. The content from the example was electrodynamics. Problems had to be handled with an intelligent tutor system while examples were provided in the form of videos of a screen-logging program that displayed the solution steps of an expert solving the problem with the tutoring system. Completeness of the examples was operationalised as either including justifications (complete) during problem solving or excluding them (incomplete). Furthermore, the learners were either instructed to paraphrase or self-explain during the example and prior to the video. This instruction was supported by voice-over prompts during the video. In contrast to other studies, the learners were first provided with a problem to be solved and then worked with one of the four video-conditions. During the 110 minutes session each student was provided with three problems, followed by one video-based example respectively. Speech during the videos was recorded as well as on-screen activity. To test for robust learning, the students completed an exam about one month after the experiment. The results show that the generation of self-explanations is superior to paraphrasing when it comes to learning. The learners who were instructed to self-explain

[9] The third condition in the experimental study by Renkl & Atkinson (2003) was independent problem solving without example.

acquired more knowledge and were better in problem solving. Those results were found to be true across different situations, assessments, and measures. Therefore the authors conclude that it is more important to generate explanations than to be provided with complete examples. Examples should thus be designed to foster self-explanation activities in the learners.

Even though gaps in worked examples have shown to be an effective means to foster learning and transfer, feedback and expert comments are important information for the learner in order to make sense of the incomplete solution presented.

The positive effects of incomplete examples have been attempted to be explained by cognitive load. Paas (1992) conducted a study comparing a computer-based training in the domain of mathematics with conventional problems, worked problems, and examples that had to be completed by the learners (partly worked-out problems). Cognitive load was measured with a mental effort rating and has become obvious to be lower during transfer for the partly and completely worked out problems compared to the conventional problems. Against his hypothesis, there were no differences of cognitive load during the learning process between the three conditions.

Another instructional approach to improve the elicitation of self-explaining activity already introduced with the study by Hausmann and VanLehn (2010) is the introduction of prompts, that have shown to appear jointly with incomplete examples.

Self-Explanation Prompts

When exposed to a certain instructional approach, not every learner applies the same strategies of working with the learning material provided. Also, the strategies applied vary in respect to their range of applicability and effectiveness. Like in the basic aspect of principle application during the learning process (see 2.1.1), hints or prompts support the learner in being more effective. Generally, research on worked examples has revealed positive effects of supporting self-explanations by prompting learners (Berthold & Renkl, 2009; Chi et al., 1989; Renkl, 2005; Schworm & Renkl, 2006). Those prompts are often 'short, attention-guiding explanations' (p. 74) that are provided either before or after the learning content is presented (Rummel et al., 2009).

As stated earlier, not all learners automatically explain the example and processes to themselves (Chi et al., 1994; Renkl et al., 1998; Roy & Chi, 2005). In some learning environments the mass of details can be overwhelming especially for novices without the additional instruction by adding self-explanation prompts (Crippen & Earl, 2007). In other cases self-explanations are produced but are not helpful because they are incorrect (Renkl, 1997).

In her study on drawing as a learning strategy for 5^{th}- and 6^{th}-graders, Van Meter (2001) compared three different conditions and a control group (no drawing). The drawing was applied as a learning strategy when reading a science text. Students who were prompted with questions after comparing their own drawing with an illustration provided by the instructor scored significantly better in the post-test than those who just constructed drawings and those who were provided with an illustration and instructed to compare it. The prompting questions were short and evolved from a general level[10] to the final question comprising the invitation to enhance the drawing[11].

There are different approaches in the literature as to how to categorise self-explanation prompts. Yeh, Chen, Hung, and Hwang (2010) compared two different prompts: the reasoning-based prompt on the one side that asked for the reason behind the course of an animation. On the other hand they implemented the predicting-based prompts that first requested a prediction of the next steps in the animation and adjacent to give a reason if the prediction was wrong. They conclude that the ideal prompts to support self-explanations must be oriented towards learner characteristics like the level of prior knowledge and should be adaptive. On the contrary, Rummel and colleagues (2009) found that in a collaborative learning environment additional prompts following distinctive model phases, combined with a phase for self-explanations, did only have a small and statistically not significant effect on the collaborative process analysis. In another study, explanation prompts even hindered application-oriented procedural knowledge acquisition (Berthold, Röder, Knörzer, Kessler, & Renkl, 2011). Those phenomena are labelled the *double-edged* effect of using prompts.

Prompts are thus, besides other designs like incomplete examples one instructional method to foster self-explanations. It remains an open question what an effective prompt has to look like to have positive effects on learning and why prompts do in some cases even hinder learning.

Besides the content of examples and the instructional support provided for skill acquisition, there is a tendency towards the embedment of examples in video-cases that are presented in screen-based learning environments. The principles that have been proven effectively can be adopted and further aspects can be investigated. Furthermore, modeling cases can be integrated in worked examples and be provided efficiently and independently from time and place.

2.3 Video-Based Examples

Worked examples are more and more implemented with computers. However, the method of video-based worked-examples is relatively new and not yet well investigated. Therefore it is only known for sure that learners are motivated by watching videos while it remains open which aspects or characteristics of the videos are responsible for that effect (Baum & Gray, 1992). To be effective, videos have to be integrated in effective instructional designs that must be chosen depending on the learning goal (Seidel, Blomberg, & Renkl, 2013). To prepare for a realistic training situation and enable the learner to make the best use of it, worked examples integrated in computer-based learning programmes have already been shown to be a sensible tool in some contexts, for example in teacher training (Hilbert, Renkl,

[10] 'In the illustration/your drawing, what are the parts of the central nervous system?' (p.140)
[11] 'Should you change anything to make your drawing more accurate?' (p.140)

Schworm, Kessler, & Reiss, 2008). But also factual knowledge can be acquired by using videos as illustrative example (Seidel et al., 2013).

In contrast to worked example research, modelling is an instructional approach that has already very early applied videos as learning material, as well as cognitive apprenticeship. Both methods share the advantage of highly authentic contexts (Ertelt, 2007). Lane and colleagues (2001) characterise the amenities of videos through 'the reality and immediacy of the emotions of the actors' (p. 305) while De Bock, Verschaffel, Jannsens, Van Dooren, and Claes (2003) restrictively complement that the authenticity or the actual degree of realism is less relevant for the appraisal of the context as realistic than 'the extent to which it succeeds in getting students involved in the problem and engage them in situational meaningful thinking and interaction' (p. 445). In general, an actual performance presented for observation can convey more information than for example a description in written or oral form, and illustrate how a schema or theory can be put into practice (Bandura, 1965). To standardise modelling performance for instruction videos can be implemented to show on the one hand a highly realistic performance and on the other hand offer convenient features like repeatability, interruptibility, and support mechanisms. Those interactive mechanisms have been investigated in order to avoid the underestimation of videos as *easy* by the learners that may lead to less investigation of effort (Ertelt, 2007).

On the other hand, realistic demonstrations can convey too many information of which some might even hinder learning. Scheiter and colleagues (2009) found that realistic visualisations (rdv) of the biological process of mitosis are inferior to schematic visualisations (sv) of the same process as far as self-evaluated difficulty and performance were concerned. Generally, the aspect of complexity has been studied deeply in the scope of cognitive load theory showing that learning with highly complex learning material or too much control for inexperienced learners leads to reduced learning (see 2.2.1.1). Still, there are findings that especially in complex and ill-structured domains like BBN where no 'prepackaged prescriptions' are possible, videos can foster deep learning and prepare for practical application (Spiro, Collins, & Ramchandran, 2007), depending on prior knowledge of the learners. Another example is collaboration in a computer-mediated setting, where observing a model has been shown to be superior to learning with a script in respect to the internalisation of the script elements (Rummel & Spada, 2005; Rummel et al., 2009) (For a detailed description of the empirical studies, see 2.2.2). This research gives hints towards enhanced efficiency of video-mediated modelling compared to learning with scripts. The next step is the investigation in how far this effect can be generalised to other contexts and non-collaborative skills.

Generally, video-clips adopted as learning material should be designed or structured according to accepted principles like the social agency theory by Mayer (2005). The theory targets social aspects of multimedia learning instead of being limited to cognitive aspects. The latter – more precisely the cognitive processing - is being targeted by the activation of social responses that lead to an augmentation of active cognitive processing which results in a better learning outcome. Principles to optimise the effects of videos used in learning environments are a) Principle of voice (words are spoken in an accent- and dialect-free human rather than a machine voice), b) Principle of personalisation (words in multimedia presentation are in conversational rather than formal style), and c) Principle of image (visibility of speaker's face is not needed on the screen) (Mayer, 2005).

But even though Lane and colleagues (2001) conclude in their review on simulation in medical education that videos are an effective tool within clinical simulations, and Rummel and Spada (2005) found positive effects on computer-mediated collaboration skills, video-mediated modelling again only can work if the learner can identify with the character (Anolli et al., 2006).

Videos can thus be an effective tool for learning from examples under some conditions. The appropriate level of complexity and support as well as the adequate model displayed are crucial to make optimal use of videos as learning material. Identification with a character seems to be fundamental when learning from a video-based modeling example. A manifestation of this relation to the model is empathic behaviour of the viewer / observer (see 2.2.2.1).

2.4 Synthesis

Based on the empirical findings presented above, consequences for the acquisition of complex skills in the academic context and open gaps in research are outlined. Additionally, aspects of effective learning tasks and interacting factors relevant to learner factors for complex skill acquisition are summarised.

Generally, learning tasks that are designed as clearly defined whole-tasks are experienced as authentic by learners. An example for this kind of task is video-based modeling worked examples. The interaction of authenticity and difficulty is yet to be tested for interactions. Another research gap to be filled is the level of support as well as the model presented in the example. In order to gain insights in the most effective design for learning the support of the learner as well as the models persona should be experimentally varied. However, the level of identification of the learner with the model is related to learning and can be measured with the level of PSI to describe the relation of the viewer to the model. Thereby the field of PSI-application could be broadened. In order to shed light on the role of personality traits on learning with modeling examples as well as further differentiate the concept of PSI and empathy, scales for the assessment of the latter can be applied, too. Other affective aspects of learning are anxiety and inhibitory effects of the learning material. Their role in the context of complex skill acquisition with modeling worked examples is still an open research question. Simulation-based motor-skill trainings already have been shown to have positive effects on affect. Whether those effects are also relevant in other domains that target complex skills and include mechanisms to report the correctness of the modelled solution steps is yet to be explored.
The learners' level of prior knowledge has been shown to be a relevant dimension in respect to the method and assessment of skill acquisition. Effects of the learners' self-explanations on complex skill acquisition are yet to be clarified. Effects of different prompts on self-explanations have not yet been tested during learning with video-based modeling worked examples. Thereby the role of self-explanations in modelling and effects of different prompts could be investigated further. Especially the relation of self-explanation-production and erroneous examples need to be investigated in more depth. Likewise, correctness of modeling worked examples and the effects of coping vs. mastery models are an open research gap.
Ultimately, the measurement of cognitive load during complex skill acquisition with video-based modeling worked examples and the relation to affective aspects during learning with different examples could help to provide a clearer picture of the concept of cognitive load.

Those empirical findings provide the framework for a study in the context of medical education that aims at closing the aforementioned research gaps. This investigation in a real *uncontrolled* learning situation in academic context will comprise those factors that have shown to affect complex skill acquisition with modeling examples to support internal validity. Also, aspects of design-based research will be continuously applied with the purpose of enhancing external validity of the empirical results.

2.4.1 Design-Based Research

In order to link research on learning and practice in education, the method of design-based research has been developed. Characteristics of this approach are (a) Research in existing educational contexts opposed to laboratory settings, (b) Assessment of the learning context and focussing on the design, (c) Inclusion of multiple methods, (d) Iterative adjustments, (e) Comprehension of practitioners from the field, and (f) Develop principles for defined contexts with the goal of enhancing the theory as well as having a practical impact (Anderson & Shattuck, 2012; McKenney & Reeves, 2013). A large field of application is research on technology-enhanced learning environments (Wang & Hannafin, 2005).

On the downside, design-based research often concentrates on the description of contexts rather than reporting effects sizes. Thereby the desired outcome of best practice rules are missing (Anderson & Shattuck, 2012).

Within this study, two phases of data collection took place, referring to the iterative character of design-based research. Before the 1st data collection started, an experienced practitioner from the field was interviewed in order to define the learning goals as well as the skills that had room for improvement. This process was complemented with a literature review aiming at the detection of problem definition as well. Those problem-analyses were then included in the development of a video-based modeling worked example for complex skill acquisition.

When working with examples (*here*: simulations), students often request repeated exposure (Paskins & Peile, 2010). This request can be met with repeatable video that also have the advantage of low costs for repetition which is a major advantage of video-based examples in contrast to the otherwise popular method of role play that cannot be repeated to the same extent (Anolli et al., 2006; Cook et al., 2010; Issenberg et al., 2005). The learners also had the opportunity to repeat distinctive scenes from the video. Subsequently, the same task had to be performed with a SP on the basis of what was taught with the modelling examples. So we can talk about a repetition even though it is with different patients and more or less realistic (via computer or face2face).

After the 1st round of data collection (1st study) the students were asked for feedback on the learning environment as well as contextual factors. This feedback, as well as the findings of the 1st study were then evaluated and integrated in the development of a 2nd round of data collection (2nd study). The same process of feedback collection took place also after the 2nd study to prepare for further developments.

For the 1st study, only part of the research questions developed from the literature review were targeted in order to avoid an overload of the learners by huge amounts of questionnaires to be answered. The focus of the 1st study is to investigate the effects of video-based modeling worked examples on the acquisition of the complex skill of BBN. The role of self-explanations as well as different mechanisms to support self-explanations during modeling worked examples will be investigated as well as the role of cognitive load and negative affect in the learning process.

3 Research Questions

3.1 RQ I: Effect of Video-Based Modeling Worked Examples on Complex Skill Acquisition

To what extent does the introduction of video-based modeling worked examples have an effect on the acquisition of a complex skill?

In the scope of RQ I the effect of abstract modelling with a video-based worked example on learning will be analysed. Performance of the complex skill of BBN is expected to be enhanced by the introduction of a video-based worked example showing a model performing the complex skill, compared to a reduced instructional approach consisting of a text-based declarative knowledge input only. In the condition learning with a video-based worked example the abstract knowledge provided in the text is translated into concrete knowledge. Worked examples have been shown to be an effective instructional approach for complex skill acquisition before (Schworm & Renkl, 2007), also in combination with modelling (Rummel & Spada, 2005; Rummel et al., 2009) and is expected to be functional in this study as well.

Hypothesis: It is thus expected that learning with the video-based modelling worked examples has a positive effect on the outcome performance compared to the text-based knowledge input.

3.2 RQ II: Effect of Self-Explanation Scaffolding on Process and Outcome of Video-Based Modeling Worked Examples

To what extent does self-explanation scaffolding with incomplete video-based modeling worked examples affect processes and outcomes of video-based modeling worked examples?

Worked examples are most effective if the level of complexity and support match with the learners level of prior knowledge (Ginns et al., 2003) and facilitate active engagement in understanding the solution presented (Crippen & Earl, 2007). This fitting of learning material and learner prerequisites is required to facilitate the learning process and support the engagement in relevant learning activities, operationalised as germane load (Kalyuga, 2011).

One aim of this study is to compare a complete to an incomplete video-based modeling worked example and investigate the effects on the learning processes as well as outcomes. To teach highly complex content in a cognitively controlled manner video-based learning material should be displayed in 'bite size chunks' (Spiro et al., 2007, p. 97). Like this it can be worked with rich cases that have many lessons to teach without having to deal with *oversimplification* (Salomon, 1981) on the learner's side. This is supported by empirical findings of positive effects of incomplete examples on transfer (Stark, 1999). While it is widely agreed upon the important role of affect during the learning process, empirical work on this matter is connection with incomplete examples is missing, except for the hypothesis that gaps should *involve* the learners more into the scenario shown (Anolli et al., 2006). Research on incomplete examples and cognitive load (Paas, 1992) has not yet shown to be conclusive.

Hypotheses: It is thus expected that learning with an incomplete video-based modelling worked example has a positive effect on the learning outcomes knowledge and performance compared to the complete example. The evaluation of the negative emotion anxiety and inhibition in the context of video-based worked examples is investigated exploratively to add insights on this matter to the field of modeling examples presented in the form of a video. There are no effects expected of the completeness of the video-based modeling worked example on cognitive load during the learning process.

3.3 RQ III: Effect of Self-Explanation Prompts on Process and Outcome of Video-Based Modeling Worked Examples

To what extent do self-explanation prompts during the gap of an incomplete video-based modeling worked example affect learning process and outcome?

Incomplete worked examples (completion principle) have been shown to support the quality of self-explanations for close and medium transfer (Shen & Tsai, 2009; Stark, 1999) which can thus be identified as one mechanism how incomplete examples support learning. But incomplete video-based worked examples should improve learning only under the condition of sufficient support. In order to further support the elicitation of self-explanations, prompts have empirically been shown to be effective (Berthold & Renkl, 2009; Chi et al., 1989; Renkl, 2005; Schworm & Renkl, 2006). Besides first hints from an empirical study by Hausmann and VanLehn (2010) it remains an open question whether this holds true for video-based worked example that can be categorised as being very complex. As only correct self-explanations have a positive effect on learning (Berthold & Renkl, 2009), positive effects can only be expected if the complexity of the examples matches the learners' level of expertise.

Another aim of the study is thus to investigate effects of different self-explanation prompts during incomplete video-based modeling worked examples on the learning process as well as on the acquisition of knowledge and performance of complex skill. The effects of an incomplete example are compared to those of two different self-explanation prompts: Activity prompt and explanation prompt, and the combination of both.

Hypotheses: It is thus expected that self-explanation prompts should be beneficial for the production of self-explanations and the investment of germane cognitive load when learning with a video-based modeling worked example.
The effect of self-explanation prompts in incomplete examples on inhibition and anxiety is investigated exploratively.
The effect of self-explanation prompts on the learning outcome will be investigated exploratively due to the inconclusive state of empirical results from the field. Positive effects

on the learning outcome are only expected under the condition that the self-explanations produced are correct.

3.4 RQ IV: Relation of Process and Outcome Variables of Video-Based Modeling Worked Examples

To what extent are cognitive load and negative emotions related to the learning outcome when learning with a video-based worked example?

Another aim of the study is to investigate the relation of germane load during working with a video-based modeling worked example and anxiety and inhibition related to BBN before and after the video-based modeling worked example to the learning outcome.

Germane cognitive load is defined as intentional cognitive effort (Kalyuga, 2011). When this is experienced by a learner it is therefore by definition positively related to the learning outcome. However, the measurement of the different concepts of cognitive load is still under negotiation. Another manifestation of cognitive effort during learning is self-explaining the problem or solution steps by the learners (Chi et al., 1989; Pirolli & Recker, 1994). The role of negative emotions in the learning process has been found relevant by numerous empirical studies, especially as anxiety accounts for large parts of variance why people avoid frightening tasks such as BBN (Merker et al., 2010). Still, there are contradicting results of the kind of effect negative emotions have on learning. On the one hand anxiety has been shown to have a positive effect on outcomes of observational learning with anxious subjects imitating more (Bauer et al., 1983; Sarason et al., 1968) while on the other hand (test) anxiety has shown to inhibit learning (Zeidner, 1998, 2007).

Hypothesis: It is thus expected that germane load is positively related to knowledge acquisition and performance of the complex skill of BBN as well as to self-explanations. The results will also give hints to the validity of the measuring instrument.
The role of anxiety and inhibition when learning with a video-based modeling worked example is investigated exploratively.

4 Method

4.1 Participants

N = 105 medical students from LMU participated in the first study. The study was conducted during the so called *SP-Week* with students studying in their 3rd or 4th year (clinical semester) in winter term 2009/10. The control condition was implemented during summer term 2009. While the experimental condition includes all data sets available, the control sample was randomly selected out of all data available. Only those students were included in the study that produced complete data sets. Incomplete data was on the one hand caused by technical problems in respect to implementation of the learning material as well as data backup. On the other hand those data sets were excluded that were incomplete because the students did not voluntarily agree to be videotaped. Finally, those students with massive linguistic problems due to German not being their maternal language were excluded from the analyses.

In the first weeks of the data collection from the experimental condition the pilot phase led to some modifications in the learning material which is why only the data sets that were collected with the final version are included.

All in all the experimental condition is composed of n = 68 datasets while the control condition comprises n = 37 cases (see Table 2). Gender of the learners allocated to the different conditions is balanced (Pearson's Chi^2=1.930, p=.587).

Table 2: Experimental design and dissemination of N = 105 cases to conditions.

<table>
<tr><th colspan="6">Video-Based Modeling Worked Examples</th></tr>
<tr><td rowspan="8">without</td><td colspan="5">with</td></tr>
<tr><td colspan="5">Self-Explanation Scaffold</td></tr>
<tr><td>without</td><td colspan="4">with</td></tr>
<tr><td rowspan="4"></td><td colspan="2" rowspan="2"></td><td colspan="2">Activity Prompt</td></tr>
<tr><td>without</td><td>with</td></tr>
<tr><td rowspan="2">Explanation Prompt</td><td>without</td><td>n = 15</td><td>n = 14</td></tr>
<tr><td>with</td><td>n = 10</td><td>n = 14</td></tr>
<tr><td>Complete condition
n = 15</td><td colspan="4">Incomplete condition
n = 53</td></tr>
<tr><td>Control condition
n = 37</td><td colspan="5">Experimental condition
n = 68</td></tr>
</table>

While the experimental condition includes all complete cases the n = 37 cases of the control condition have been selected randomly from the whole data (video-recordings) available by picking videos from different weeks and including about the same number of male (n = 18) and female (n = 19) students. To give a realistic picture of the effects of the independent factors on performance and correlations of dependent variables with the complex skill of BBN, not all data available is included but only of those learners the performance could be rated.

Most medical students of the experimental condition that took part in the course during winter term 2009/10 participated during their 4th year (88.7%) of studies. Seven students (11.3%) still were in their 3rd year, and for six participants the information is missing. Gender is nearly balanced with 51.5% (35) females and 48.5% (33) male students. The mean age is 25 years with a standard deviation of four years and seven month (information missing for six participants) and a range from 21 to 52 years. The students were asked on a questionnaire prior to the video-based modeling worked example whether they already took part in Modul 2 which is a course that teaches general behaviour in difficult communication situations (German item: *'Ich habe Modul 2 bereits besucht'*). The majority of 85% indicated that they had prior experience in the field of difficult communication situations. Another aspect of prior experience collected was if they already delivered bad news (German item: *‚Ich habe bereits eine schwierige Nachricht übermittelt.'*) which 70% answered with yes. On the contrary, the concept of the spikes-steps was only familiar to 21% of the students from the experimental condition (German item: *‚Hatten Sie den Text vorher (d.h. vor der heutigen Lernsitzung) bereits gelesen?'*). The students from the experimental condition were randomly assigned to different experimental conditions that will be described in more detail in 4.3 Experimental Conditions.

For the participants of the control condition information on age, semester, and prior experience is missing as the study only started after they were available for data provision[12]. To check for comparability between the conditions, crosstables have been generated comparing the control group and the treatment group in respect to the distribution of the student's gender, no significant differences can be reported (Pearson's Chi^2=.000, p=.991).

4.2 Learning Environment and Procedure

The data collection for the study was embedded in the introductory session *Breaking Bad News – Einführung in die Übung "schwieriges Gespräch"* which took place in the scope of the *SP-Woche Chirurgie* which translates to *Simulation* and *Examination* (*Prüfung*) that is part of modul 3[13] (surgical topics) of the medical curriculum (mecum) at the LMU (For the plan of the whole week see Appendix 1: Plan for SP week with introduction to breaking bad news.). The targeted skill of the session that takes place on day one (Monday morning) is BBN to a

[12] The videos-taped performances by the students from the control group cannot be traced back to individuals as they have been recorded without mentioning the students' names.

[13] The curriculum has changed after the data collection and modul 3 has been merged with the internistic orientated modul 2 to one modul 23.

standardised patient (SP) - more exactly delivering a cancer diagnosis – which is performed and captured on video on day 2 (Tuesday afternoon). The study has been conducted in close collaboration of the chair of educational psychology and empirical pedagogy with the surgical clinic, both LMU. Each week of the semester another group of 16 to 20 students participated in the *SP-Woche*.

The following description of the learning environment refers to the experimental condition implemented in winter term 2009/10. Generally, the topics addressed during the learning intervention were chosen in response to the results of a needs assessment that examined videos of BBN conversations, analysed literature from the field, and took into account an interview with an experienced physician working as surgeon (who practices as well as teaches students). For a plan of the course with time schedule, see Table 3.

Table 3: Time and workflow schedule.

			time		content
Day one (Monday)			10 min	09:30-09:40	Short introduction in the project
			15 min	09:40-09:55	Text on the spikes
			2 min	09:55-09:57	Briefing for the simulation
	Online learning environment	Phase 1	10 min	09:57-10:07	Application-oriented knowledge test (pre)
			10 min	10:07-10:17	Factual knowledge test (pre) + four items on affect
		Phase2	30 min	10:22-10:52	Video-based modeling worked example
		Phase3	10 min	10:52-11:02	Application-oriented knowledge test (post) + four items on affect + 1 item on cognitive load
			10 min	11:02-11:12	Factual knowledge test (post)
Day two (Tuesday)			20 min		Breaking bad news to a standardised patient

To standardise time on task the time span the learners had for the different tasks was pre-defined. After a short introduction into the field, given by the instructor, the students were asked to read the text on the spikes-protocol (see Appendix 2: Short version of the spikes text based on Baile and colleagues (2000).) as well as a short briefing for the simulation on the next day (scenario) where the problem is described (see Appendix 3: Briefing for the simulation.).

The learning materials are based on the spikes protocol that formalises the procedural skill of BBN. Baile and colleagues (2000) developed the protocol for difficult communication when delivering bad news that comprises six steps. The spikes can increase the physicians' confidence as shown when applied in workshops (Baile et al., 1997) and perceived competence (Bonnaud-Antignac, Campion, Pottier, & Supiot, 2010). It is also possible to avoid the effect of keeping mum about negative information to a certain extent. Rosen and Tesser (1970) were the first to label the reduced probability to communicate negative information in contrast to positive information as MUM effect (for more information on the MUM effect, see 2.1.2). Despite preventing the MUM effect, Baile and colleagues (2000) had four more goals in mind to tackle with the spikes and thereby enhance communication in difficult situations: a) gather information from the patient, b) transmit medical information, c) provide support for the patient, and d) determine if the patient wishes to collaborate in strategy and treatment plan. Those aims are pursued by the application of six behavioural steps proposed by the protocol to guide the doctor through the conversation. Each of the six letters of the spikes stands for one step that contains at least two sub steps. S – Setting up the interview, P – Addressing the patient's perception, I – Obtaining the patient's invitation, K – Giving knowledge and information to the patient, E – Addressing the patient's emotion with empathic responses, and S – Strategy and summary.

The spikes text was provided to the learners in English and defines how good communication in difficult situations is supposed to take place. Even though the perfect strategy for communication in a partly unknown situation cannot be foreseen (Realdon et al., 2006), the six steps offer a multitude of practical *if-then* rules of behaviour that should generally fit into the previously defined context and give a clear definition of the highly complex skill as recommended in the review by Salas and colleagues (2009). Obviously, the six principles cannot easily be copied neither to the worked examples cases the learners have to work with nor to the conversation they are having with the standardised patients. Instead, the underlying principles have to be recognized and transferred to the new cases (VanLehn, 1996). The briefing is relevant for the learners to understand the problem situation displayed in the video-based worked example, because the student in the video received the same scenario before he performed the BBN. This structure is in accordance with the timing principle by Shen and Tsai (2009).

Figure 1: Setting of the data collection with investigator in the front and each student sitting behind one computer screen.

After 17 minutes all students were supposed to have finalized reading the texts and were asked to login their own computer (Figure 1) and enter the first phase (pre-test). Each participant was provided with an individual code and had one computer to work with. When logged in, a video-based knowledge pre-test appeared on the screen. The students were instructed to answer the three questions without looking into the spikes text. Subsequently, a factual knowledge test consisting of four questions about the spikes started (for the description of both knowledge tests, see 4.4.1), followed by four items on the learners' affective state (4.4.4). Both tests had to be completed within a maximum of ten minutes, respectively.

In the second phase (worked example) of the learning the students watched the video-based worked example and worked with it (see 4.2.1) for a maximum of thirty minutes. During this phase it was explicitly wanted that the students worked with the spikes text. After the time was over, everyone was asked to stop the activity and start with the third phase (post-test) which structured accordingly to the pre-test but included one additional item on cognitive load. Again, the knowledge tests had to be handled without the spikes text.

On the next day, the students returned for the delivery of a cancer diagnosis to a standardised patient (see 4.4.2) in order to test their complex skill of BBN. Each student met with one patient alone in a room for a maximum of twenty minutes and was video-taped by the investigator.

4.2.1 Video-Based Modeling Worked Example

The video-based worked example is based on one of the modeling cases from the control group. It was selected because the BBN performed contains both correct performance steps of the targeted behaviour and incorrect or suboptimal actions. It can therefore be categorised as displaying a realistic average performance of the complex skill. This is reflected by the number of spikes-steps that are performed by the doctor in the video: 17 steps have been coded which is consistent with the mean number of spikes-steps performed in the control group (*Mean* = 17.82, *SD* = 2.735). For an overview of the number of spikes-steps performed in the video-case see Table 4. In order to avoid mix-up of the first and last spikes-step which are both labelled with '*S*', the second S, describing the step 'Summary and Conclusion' will be marked as '*Z*' instead (for a description of the Spikes, see Appendix 2: Short version of the spikes text based on Baile and colleagues (2000).).

The basis for the worked example is a modelled video-case. As described in the theory part, rule-oriented learning points containing the principles behind the presented behaviour have a positive effect on reproduction and generalisation of the modelled behaviour (Decker, 1984), which has been shown to be true in different learning environments (Rummel et al., 2009). Therefore the video-based modeling worked example is structured according to the spikes steps the students read about in the introductory text.

The design of the video-based example used in this study complies with the design principles summarised by Shen and Tsai (2009). Only the fading principle has not been realised out of time constraints.

Table 4: Number of spikes-steps performed in video-case A.

		Maximum of steps per spikes
S	5	8
P	1	3
I	0	2
K	4	5
E	2	5
Z	5	8
Total number of spikes-steps	17	31

The video-based example was divided into eleven short episodes. It has been shown that enhanced feedback has positive effects in the use of virtual patients (Cook et al., 2010) and is one of the success factors during simulation defined by Issenberg and colleagues (2005). Therefore, an expert's comment was shown in all conditions after an important spikes-step occurred or was missing in the video-based worked example, highlighting correct and incorrect aspects of the solution as well as making explicit the *why* and *how* as described in the process principle by Shen and Tsai (2009). The video could be paused by the learners. After watching all episodes they were also able to get back to single episodes independently of the actual order. Total time on the worked example was limited to 30 minutes, including the duration of the video-case itself with 11 minutes and 16 seconds.

4.3 Experimental Conditions

The instructional approach applied in this work is based on a combination of worked examples and modelling. The learning material consists of a video-based modeling worked example. In this round of data collection, cases are presented in the format of video-based modeling worked-examples and vary in terms of instructional support. By using a video-case for the example, the learning material can be categorised as very realistic and integrated. The aspect of realistic learning material is in accordance with the presentation principle and the media principle by Shen and Tsai (2009), as well as empirical results from different areas of research like observational learning, video-research, and investigation of PSI. As the learners' prior knowledge is estimated as being relatively low in respect to the complex skill of BBN, the task was to study the given example instead of imagining the procedures and the underlying concept (imagination principle) (Ginns et al., 2003; Shen & Tsai, 2009).

Two instructional approaches have been applied to support the acquisition of the complex skill of BBN by self-explaining: Incomplete worked-examples and self-explanation prompts.

As the example includes correct and incorrect behaviour, self-explanations should yield the best possible learning effects (Siegler, 2002).
Already a traditional worked example can be perceived as very complex. In the case of a video-based modeling worked example the effect can be assumed to be even more relevant. A reason why this is not conducive for novice learners is the overwhelming number of details presented. Inefficient behaviours like means-end search can be avoided with hints (Spiro et al., 2007). Those hints may be prompts that guide the learners' attention to important aspects (Rummel et al., 2009), thereby support self-explanations and follow the self-explanation principle (Crippen & Earl, 2007; Shen & Tsai, 2009). During the study different kinds of prompts to support the students' self-explanations were provided to those learners who received an incomplete worked example. But one student always received the same prompts during each of the four gaps of the video-based modeling worked example (A).
As a basis, every gap contained the request to answer how the doctor should behave now (original German formulation: *'Wie sollte der Arzt sich jetzt verhalten?'*). To further investigate what kind of prompt best supports correct self-explanation, this basic request was experimentally complemented with two more specific prompts. One prompt to foster self-explanations directed towards an elaboration on the basis of the spikes-steps (referred to as *activity prompt*) and another prompt to self-explain in more depth (referred to as *explanation prompt*). Especially the latter explanation prompt was implemented to support causal argumentation.

The same video-case is used as a basis for the example in the different conditions. The participants of this study all worked with the same video-case, in the following referred to as video-case A. The between subject factor is the structure of the examples varying in completeness and self-explanation prompt.

Control Group

N = 37 students from the semester prior to the introduction of the intervention (summer term 2009) serve as a control condition. The students took part in the same course as the experimental groups but did not receive any training with a learning environment. Instead, they were given a very short introduction to the topic of BBN in general. However, they were provided with the same briefing for the simulation (see 4.2) as the experimental conditions and also had to read a text on the spikes. Instead of the short version the students from the experimental condition received, the student from the control condition were asked to read the full paper by Baile and colleagues (2000) on the spikes-protocol. This means they had the same (theoretical) information about their task to be performed on the next day with a standardised patient but not in a condensed form as the students from the experimental condition. Phase one to three of the time schedule (Table 3) were not implemented in the control condition. Just as in the experimental condition, the students from summer term 2009 met a standardised patient on day two (Tuesday) and had twenty minutes for BBN as well as receiving feedback from the patient. Those conversations were videotaped on a voluntary basis. Six different SPs were involved in the data collection during summer term 2009 with four being female and two male.

Complete Video-Based Modeling Worked Example

The students from the experimental conditions were working with the online learning environment[14] described in 4.2 with the main part consisting of a video-based modeling worked example. N = 15 randomly assigned participants worked with the complete video-based worked example. In this condition phase two of the online learning environment provided an example that was shown entirely to the students. After each of the eleven episodes (see 4.2.1) the video was paused and an expert's comment appeared automatically on the right hand side of the freeze image, see Figure 2. After reading the comment the learners could continue with the next scene by clicking on a blue arrow. This group completed day one with the learning environment as described in Table 3, followed by BBN to a standardised patient on day two. As for the control group the performances of BBN were videotaped if the students approved.

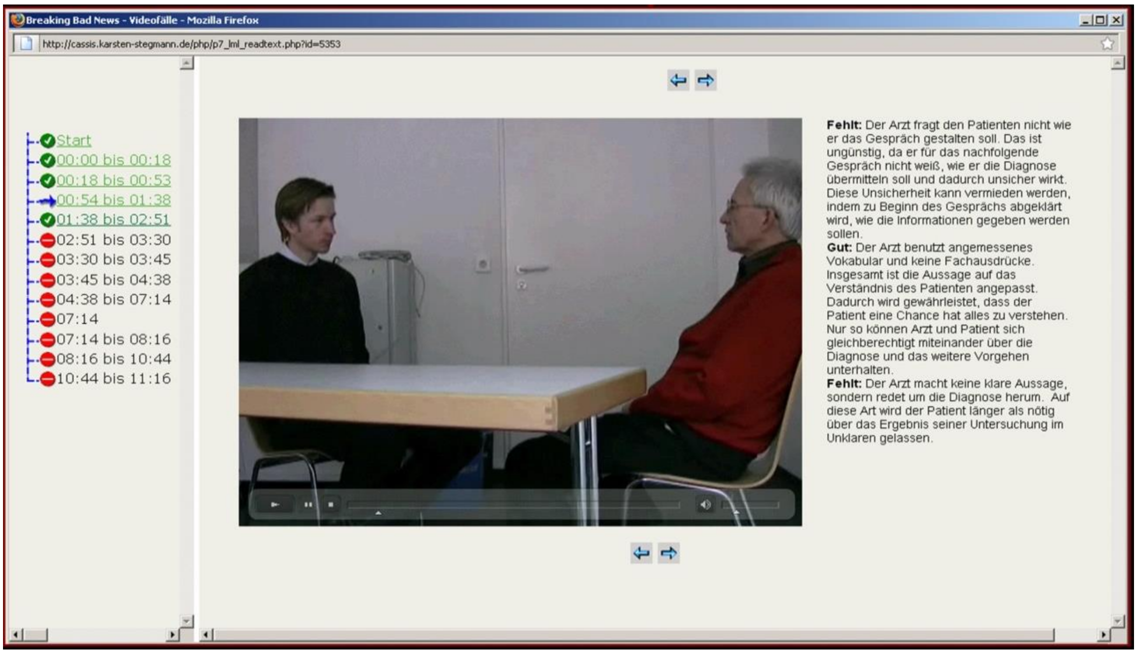

Figure 2: Screen-shot of the video-based modeling worked example A with an expert's comment.

Incomplete Video-Based Modeling Worked Example

The first factor to vary within the experimental condition is the completeness of the video-based modeling worked example. One randomly assigned group of N = 15 students passed the whole process of the learning environment with the three phases as well as meeting a standardised patient on day two as did the students from the complete condition. They also worked with the same example the learners from the complete were working with, but the example included four gaps. During those four interruptions, the learners were first shown a blank text field on the right hand side of the video scaffolding the learners to answer the question 'How should the doctor behave now?' (German: ‚*Wie sollte der Arzt sich jetzt verhalten?*'). The expert's comment only appeared on screen after the gap. During the respective four gaps in the video-based worked example time for typing in self-explanations was limited to three minutes. After this time span the screen changed from the text field for

[14]The learning environment was hosted on cassis.karsten-stegmann.de/mecum.

the self-explanations to the expert's comment. If the learners had typed in text, it was shown above the comment for comparison (Figure 3).

Of the 68 cases included in the experimental group, 22% (15 students) received a complete worked example while the remaining 78% (53 students) worked with incomplete examples. The skewed distribution to complete and incomplete examples is due to the second independent factor: during the gaps, different self-explanation prompting conditions were varied.

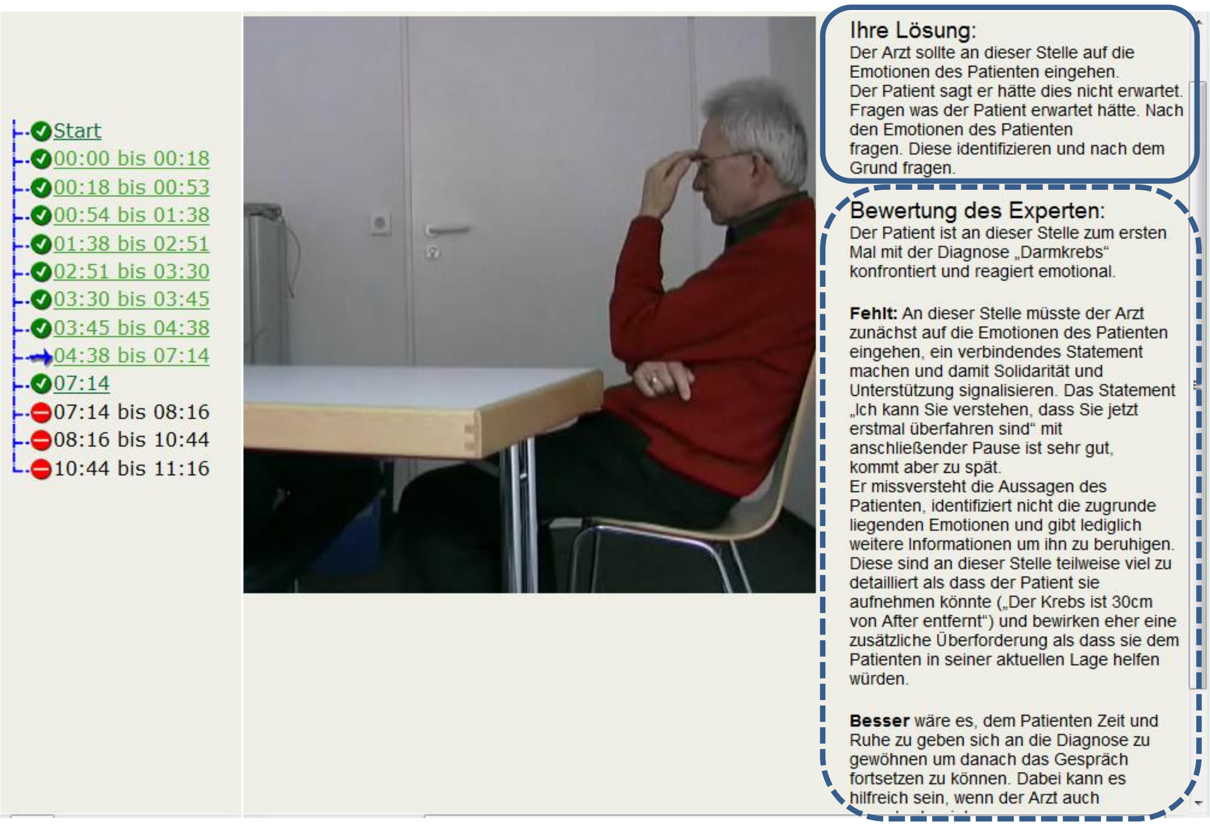

Figure 3: Screenshot of the incomplete modeling worked example condition with a learner's answer (encircled) and the expert's comment (dotted line).

Incomplete Video-Based Modeling Worked Example with Activity Prompt

N = 14 participants of those students who worked with an incomplete modeling worked and passed the whole process of the learning environment with the three phases as well as meeting a standardised patient on day two, were provided with prompts to scaffold self-explanations. In this condition the learners were scaffolded in each of the gaps in addition to the question 'How should the doctor behave now?' with the activity prompt ‚Please, refer to the specific spikes-steps!' (German: ‚Beziehen Sie sich dabei bitte direkt auf spezifische Schritte der Spikes.'), see Figure 4.

Incomplete Video-Based Modeling Worked Example with Explanation Prompt

N = 10 participants of those students who worked with an incomplete modeling worked and passed the whole process of the learning environment with the three phases as well as meeting a standardised patient on day two, were provided with prompts to scaffold self-explanations. In this condition the learners were scaffolded in each of the gaps in addition to the question 'How should the doctor behave now?' with the explanation prompt ‚Please, explain why the doctor should behave this way in your opinion!' (German: ‚Erklären Sie bitte warum sich der Arzt Ihrer Meinung nach so verhalten sollte!'), see Figure 5.

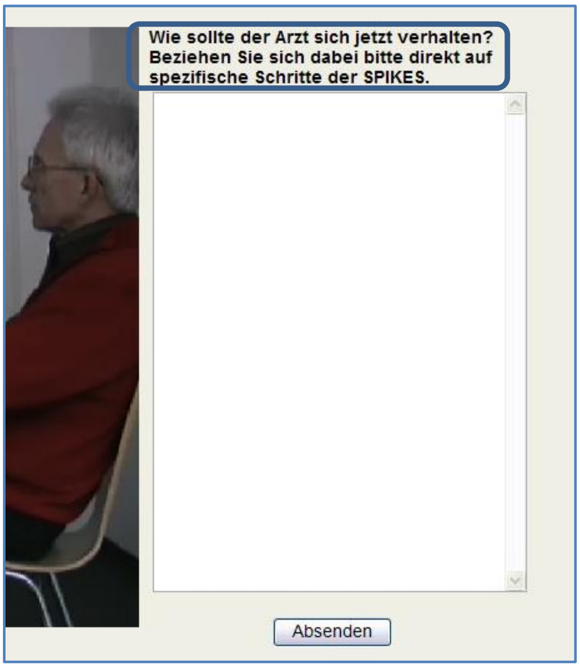

Figure 4: Section from screenshot of the activity prompt (encircled) condition in the incomplete modeling worked example.

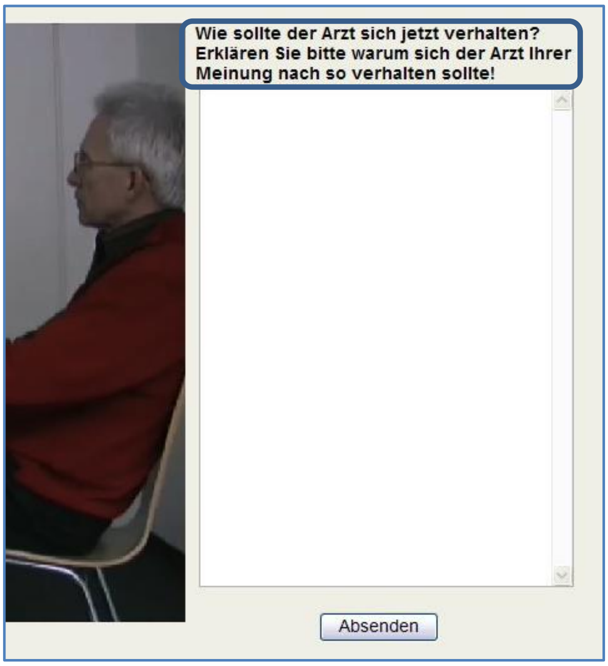

Figure 5: Section from screenshot of the explanation prompt (encircled) condition in the incomplete modeling worked example.

Incomplete Video-Based Modeling Worked Example with Activity Prompt and Explanation Prompt

N = 14 participants of those students who worked with an incomplete modeling worked and passed the whole process of the learning environment with the three phases as well as meeting a standardised patient on day two, were provided with both prompts to scaffold self-explanations. In this condition the learners were scaffolded in each of the gaps in addition to the question 'How should the doctor behave now?' with the activity prompt as well as the explanation prompt: , Please, refer to the specific spikes-steps! Please, explain why the doctor should behave this way in your opinion!'.

4.4 Dependent Variables and Data Sources

The complex skill of BBN consists of complex social behaviour and is shown in a complex social setting. Therefore the method of data collection and data analysis cannot be reduced to a minimum of complexity but has to include several relevant factors. According to Salomon (1992) the most important main effects in a study of human learning and behaviour in the real world are interactions (with context factors): '*To understand learning [in such innovative classrooms] one needs to study the whole learning environment and the way it and the individuals in it interactively change (p. 171).*'

It is furthermore important to differentiate between learning as knowledge acquisition and the performance thereof (Bandura, 1965). In our setting the BBN performance is the key outcome variable. The performance measures may differentiate from the knowledge the learners acquire, operationalized with knowledge tests.
Therefore, different measurements of factual and application-oriented knowledge on the spikes as well as the complex skill of performing a BBN conversation with a Standardised Patient are applied. While the knowledge tests are computer based questionnaires with open questions to be answered by typing in the solution (4.4.1), the performance of BBN is embedded in a simulated doctor-patient setting that is video-taped for analyses of the behaviour shown by the learner (4.4.2).

Figure 6: Exemplary picture of a learner filling in her answers in a screen-based questionnaire.

Besides measurements of learning outcomes, variables expected to impact on the learning process were collected with screen-based questionnaires (see Figure 6) before, during, and after the video-based modeling worked examples was processed. Before and after the example, anxiety and inhibition were assessed with a questionnaire as well as cognitive load during the examples (4.4.4). As was the latter, self-explanations were produced during the gaps of the incomplete examples but not in form of a questionnaire but with free-text fields (4.4.3).

4.4.1 Knowledge Tests

To get a valid picture of the students' prior knowledge on the spikes as well as their knowledge acquisition during the learning intervention, two different measurements were applied. On the one hand the learners received a traditional factual knowledge test with open questions in order to assess their factual knowledge on how to break bad news. On the other hand a video-based test was introduced to assess procedural knowledge on the application of the spikes while delivering a diagnosis. Data on both forms of knowledge was collected in phase one and three of the learning environment (see Table 3), allowing for inferences on status prior to the intervention and after it.

Prior Factual Knowledge – During phase one of the learning environment, prior to the video-based modeling worked example, all students were asked to answer a computer-based knowledge test comprising four open questions on the spikes steps S, P, E and EQ. An exemplary question is question number two targeting spikes step P *'Please elaborate how to enact during step "P – assessing the patients perception". Which subsidiary steps does this comprise?'* (for the full questionnaire with the German items see Appendix 4: Factual knowledge test including four questions.). Those have been selected in collaboration with a medical expert and by analysing videos of prior semesters because the medical students had shown problems in applying them. EQ is part of the spikes-steps E as well as Z and relates to the actual empathic statements the participants make during BBN. The time span given to complete the four questions was limited to ten minutes and the coding scheme is based on the SPIKES by Baile and colleagues (2000). In total, learners could reach a maximum of 17 points due to sub-aspects of the spikes steps. The ratings of the data collection range from a minimum of 0 points to a maximum of 16 points reached for the pre-test (M = 8.227; SD = 2.708). All answers were coded by one to two trained coders. Prior to the coding, every coder completed a coding training until reaching similarity values with the gold standard of every of the 15[15] categories of $\kappa > .6$ which is in accordance of the conventions by Landis and Koch (1977) who defined Cohen's kappa of > .6 as indicating substantial accordance. The gold standard was defined by project leader and partly adapted in negotiation with student helpers.

Factual Knowledge Post - During phase three of the learning environment, following the video-based modeling worked example, the learners were again asked to fill in the four questions of the factual knowledge test. The questions were identical in formulation as well as order to the pre-test. The ratings of the data collection range from a minimum of 0 points to a maximum of 14 points for the post-test (M = 7.764; SD = 3.144).

[15] As the fourth question of the factual knowledge test asked for three examples of one spikes-step, there are only 15 different categories while 17 points could be reached.

Application-Oriented Knowledge – The application-oriented procedural knowledge test used video-taped BBN conversations[16] during which a student had to deliver a cancer diagnosis to a standardised patient: Pre- and post-test are two cases that differ in respect to the model but have the same general scenario (see description in 4.2).

Figure 7: Screenshots of the videos shown in the application-oriented knowledge test (left: E; right: Y).

Both conversations are led by a male medical student (differs between the videos) and a patient (see Figure 7): In case E it is an elderly male patient[17] while the student in case Y delivers the diagnosis to a young female patient. The videos are reduced to three scenes after which the learner is asked to write down how the doctor in the given scene should go on (for the exact questions see Appendix 5: Application-Oriented Knowledge test: Formulation of the questions following a short video clip.) respectively. Time on task was controlled by allowing the students to spend up to three minutes for the answer after each scene. The order of the two video cases is varied randomly resulting in an equal distribution of participants working with the young woman's case as pre- and post-test (see Figure 8). The targeted steps of the spikes scheme and the order of appearance were the same for the two video cases with the coding scheme is based on the SPIKES by Baile and colleagues (2000). In total, learners could reach a maximum of 15 points due to sub-aspects of the spikes steps. The ratings of the data collection range from a minimum of zero points to a maximum of ten points reached for working with video-case E (M = 3.55; SD = 1.705) and maximum seven points for working with video-case Y (M = 2.89; SD = 1.480). All answers were coded by one to two trained coders with a minimum inter-rater reliability of $\kappa > .6$ for the different categories. Prior to the final coding, every coder completed a coding training until reaching similarity values with the gold standard of every of the 15 categories of $\kappa > .6$ which is in accordance of the conventions by Landis and Koch (1977). The gold standard was defined by project leader and partly adapted in negotiation with student helpers.

This kind of test may evoke the imagination effect described in van Gog and Rummel (2010). As the prior knowledge of the learners is assumed to be relatively low the imagination

[16] Both videos are taken from the already existing sample of taped conversations of the same course before the treatment.

[17] The patient in the video-case for the knowledge test is the same as in the video-based worked example.

principle (Shen & Tsai, 2009) indicates not to work with imagining but asking the learners to study the example.

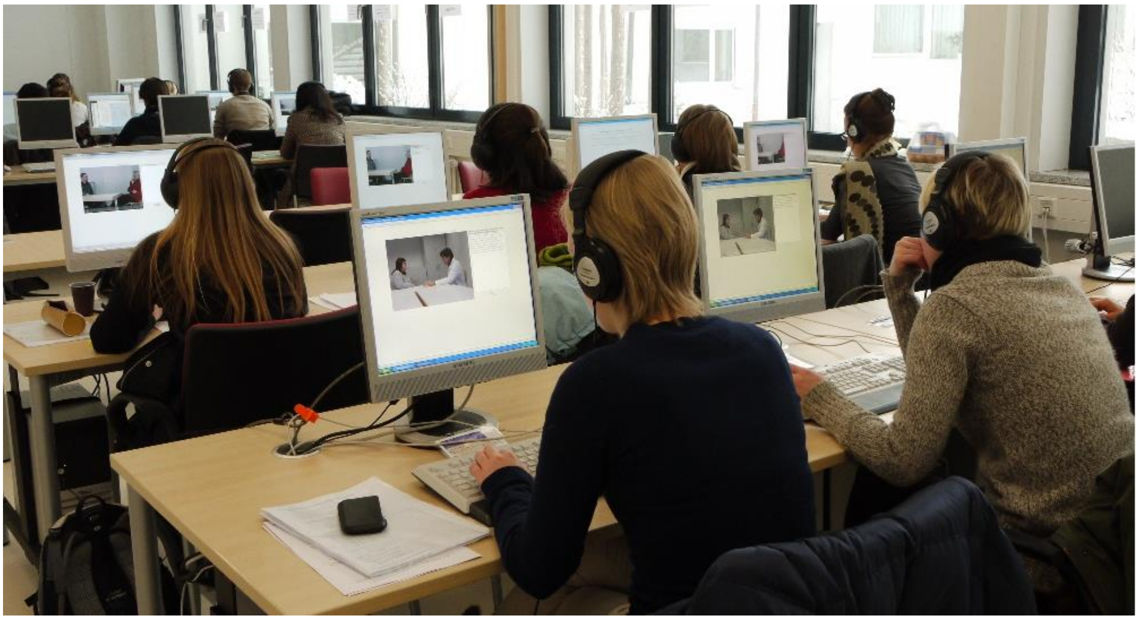

Figure 8: Students working with one of the two video-cases E and Y to assess knowledge on the spikes.

Both video-cases of the procedural knowledge test have been selected to provide insights into a variety of realistic BBN scenarios with different doctors and patients. As a side effect the cases are not completely similar. However, the order of the application-oriented knowledge test was balanced in each condition. A and a t-test showed no significant effect ($t(44) = -.135$, $p = .893$, $d = -0.04$) of the order of the video-cases within the knowledge test on the mean gain in knowledge. To avoid confounding effects of the differences of the two video-cases on the knowledge acquisition scores, the two tests have been z-standardised separately[18].

4.4.2 Video

Studies have shown that medical students do not rate their skill level in an adequate way (Evans, Leeson, & Petrie, 2007) Therefore an objective measurement is needed to get an adequate picture of the future doctor's skill. In general it is still difficult to show that simulations can improve clinical outcomes (Wood, 2010). With the help of a realistic scenario including standardised patients the complex skill of communicating in BBN situations can at least be approached.

The main goal of the training is to enhance the BBN communication the doctor has with the patient. This complex communication skill is an activity that can only be measured via the activity itself like every other competence (Kaufhold, 2006). The complex skill of communicating in a BBN situation is measured through authentic assessment. This relatively new form of assessment is applied to measures competencies or complex skills that are relevant for successful job performance (Gulikers et al., 2008). Thereby a directly link between the learning and assessment phase and the skills learners need to perform

[18] The values of video-case A is z-standardised separately from the values of video-case B.

afterwards in the working phase can be made (Gulikers, Bastiaens, & Kirschner, 2004). The closer to the future professional practice the learners perceive the assessment the deeper the learning (Gulikers et al., 2008).

The complex skill of BBN was measured by one situation with a standardised patient per student on the day after working with the video-based modeling worked example. In total, 105 data sets include a performance coding, comprising video-data of 37 coded cases of BBN conversations from the control condition (summer term 2009) and 68 from the experimental condition (winter term 2009/10) who worked with the video-based modeling worked example. In the experimental condition, seven SPs have been involved that have been randomly allocated to the students. Two male and five female SPs meet with students. Only two of the SPs involved in the control condition also participated in the treatment group, one of them being the male patient displayed in the video-based modeling worked example. For the setting of the BBN performances see Figure 9 displaying two different standardised patients.

Figure 9: Setting of the BBN conversations to measure performance with two different female SPs.

In both conditions, every student had a maximum of 20 minutes for breaking the bad news to the SP and receiving feedback from him/her. The students as well as the SPs were instructed to use about half of the time for delivering the diagnosis and half of the time for feedback. To make the setting as realistic as possible neither were the students stopped after ten minutes to start with the feedback nor were they asked to make use of the full first half for the BBN communication if they stopped earlier. This caused on the one hand some variance in duration of the conversations. On the other hand the mean time spent on BBN was a little over nine minutes (M = 9.34 minutes, SD = 2.47). Furthermore, validity is supported and bivariate correlations of duration of BBN and performance ratings showed no significant connection neither for the control condition (N = 37, r = .229, p = .173), nor the experimental condition (N = 68, r = .072, p = .558) and thereby supports the procedure of missing control of time on task for this method.

During the session no other student was in the room. The only person besides the SP was a student helper in charge of the experiment to organise the recording.

To avoid general denial from the side of the students it was voluntary to be videotaped during the session with the SP. Video is an ideal medium for studying conversations because it comprehends oral and visual reactions of both communicating parties.
A crucial factor of authentic assessment is the explicit definition of the complex skill that is targeted (Salas et al., 2009). The skill to break bad news has been operationalised based on the spikes-protocol by Baile and colleagues (2000) by operationalising the steps described in the paper, resulting in a coding scheme comprising 31 categories (for the coding scheme see Appendix 6: Coding instructions for the performance of breaking bad news based on the SPIKES by Baile and colleagues (2000).). For the analyses, the complete BBN conversations were analysed with videograph® by time-sampling. Every five seconds a category had to be coded. In addition to the 31 categories indicating the specific spikes steps two more coding-categories were possible: ‘patient talks’ and ‘no spikes step’. The codings of the respective videos were then exported to SPSS wherein the number of codings within the categories were recoded into *yes, coded* vs. *no, not coded*. Therefore the highest score for the complex skill of BBN corresponds with 31 points to the number of categories in the coding scheme.
All videos are coded by two to four trained coders with a minimum inter-rater reliability of $\kappa > .6$ for the different categories. In the case of differences between the codings the coders had to discuss it and agree on one solution. Therefore, objectivity of analysis - measured by correlation between different coders – is high. Prior to the coding, every coder completed a coding training until reaching similarity values with the gold standard of every of the 31 categories of $\kappa > .6$ which is in accordance of the conventions by Landis and Koch (1977).

4.4.3 Gaps

During four gaps of the worked example in the incomplete condition the students were prompted to type their self-explanation on the displayed scene in an empty text field. They were given three minutes to type in their answer. After this, the field to type in the explanation disappeared and the expert’s comment appeared. The self-explanations were coded on a nominal scale to either be a self-explanation or not (yes vs. no). Thereby every student could reach a maximum of four points because they were asked to give a self-explanation four times during the video. The rationale for the rating is based on the work done by Lenhard (2009), Aderhold (2008), Stark (2000), and Chi and colleagues (1989). No differentiation was made between different kinds of self-explanations.
The rating self-explanation ‘yes’ was given if the students’ answer was a) the deduction of new facts out of something already known (Stark, 2000) (e.g. ‘Give the patient time to become aware of the meaning of his words’), b) prediction or statement about future events (Lenhard, 2009) (e.g. ‘If the patient doesn’t feel well he might like to have a loved one to be present’), c) statement about conditions and consequences (Aderhold, 2008) (e.g. ‘It is important for the doctor to pause in order to give the patient time to become aware of the diagnosis’), or d) explanation of a situation or circumstances (Chi et al., 1989) (e.g. ‘Because the patient doesn’t expect the diagnosis the doctor should approach it very cautiously’). In general, a causal or conditional conjunction had to be given (e.g. because, in order to, if, …) as well as – content-wise – a doctor-patient interaction.

In a second step, a rating of whether the self-explanation is correct or not (yes vs. no) was made. The rating of accuracy was based on the accordance (content based) with the expert's comment following. The rating self-explanation 'no' was given for all other answers, especially explanations only targeting the course of disease, the superficial cognitive elaboration of examples (Stark, 2000), and reciting or paraphrasing the spikes-steps. As for the performance ratings, the text-input during the gaps was coded by one to two coders who received a coding-training until reaching a minimum inter-rater reliability of $\kappa > .6$ for the different gaps. As for the video-codings the coders had to discuss deviations in estimates and agree on one solution. Therefore, objectivity of analysis - measured by correlation between different coders – is high.

4.4.4 Questionnaire

Affect

Four items targeting anxiety and inhibition have been developed by Siebeck and colleagues (2011). Learners were asked prior and after the video-based worked example to rate on a five-point Likert scale ranging from 1: 'Do not agree' to 5: 'Agree'. For the items as they appeared in the learning environment, see Appendix 7: Items of the scale to measure affect related to BBN. Sufficient reliability of the scale with Cronbach's alpha ranging between α = .74 (pre) and α = .753 (post) for the two measuring times. Anxiety was measured with the item 'I am afraid of breaking bad news' while an exemplary negatively formulated item to measure inhibition is 'In my opinion breaking bad news can easily be done'. The mean value of the complete scale for the experimental condition is 14.23 (*SD* = 2.82; *min* = 7; *max* = 20) for affect prior to the modeling worked example and 13.83 (*SD* = 2.85; *min* = 7; *max* = 20) for affect after the example.

Cognitive Load

As there is no consistent empirical basis on how to measure cognitive load, the selection of instrument was based on considerations to limit the number of items to be answered by the learners. Difficulty experienced by the learners has been found to be closely related to the concept of germane load: The less difficult learning is experienced, the more capacity learners have to actively process the content (DeLeeuw & Mayer, 2008).

Therefore, one retrospective difficulty rating item on the learning environment and one item on the difficulty to read the spikes text have been measured to cover the concept of cognitive load (for both items see Appendix 8: Difficulty rating of reading the spikes text and working with the video-based modeling worked example.). The items were rated on a nine-point Likert scale ranging from 1 (extremely easy) to 9 (extremely difficult). Mean value for the experimental condition is 3.63 (*SD* = 1.32; *min* = 1; *max* = 6) for germane load of the learning environment and 3.52 (*SD* = 1.55; *min* = 1; *max* = 9) for difficulty of reading the spikes text.

4.5 Statistical Analyses

To analyse effects of the independent variables on the outcome measurements the following values to categorise effect sizes are applied.

Eta² (η^2) is used as a measurement of the amount of total variance of the dependent measurement that can be explained by the factor with values range from 0-1. According to Cohen (1988) the effect is categorized as small if values of η^2 = 0.01-0.06 are observed (1-6% of variance explained). Moderate effects are found if η^2 = 0.06-0.14 while $\eta^2 \geq 0.14$ describes large effects. In multifactorial designs *partial* η^2 is reported. It can be added to more than 100% because the effects of the single factors can overlap and though be included more than once which is why the non-partial η^2 will be favoured if possible.

The value of *Cohen's d* indicates the distance of two distributions by measuring their shared standard deviation. The effect is small if $d = 0.2$ while values starting at $d = 0.5$ indicate moderate effects and $d = 0.8$ point towards large effects.

The correlation r of two variables is classified by Cohen (1988) as indicating weak connections with r = +/-.10. Correlations are medium with values of r = +/-.30 and strong with r = +/-.50.

5 Results

In the following section, the research questions will be targeted by statistical analyses. Prior to that distinctive features of the dataset will be reported as preliminary results.

5.1 Preliminary Results

The measuring instruments have been checked for unintentional effects. The factual knowledge test has been shown to have lower mean values of the post-test M = 7.8 (SD = 3.36; min = 0; max = 14) compared to the pre-test values M = 8.56 (SD = 2.89; min = 0; max = 16). This is presumably due to a motivational decrease as the learners had to answer the same questions after only a short period of time (phase 1 and 3). Therefore, the analyses of the research questions will be based on the application-oriented knowledge data instead.

Due to time constraints on the side of the SPs, technical problems with the video recording and missing data from some participants, the distribution of SPs to test condition is not totally balanced for the study. In only ten (15.2%) out of 66 of the cases of the experimental condition the student met the same SP for the conversation previously seen in the video. This is not equally distributed to the conditions (Pearson's Chi^2=22.477, p=.000) with most of the learners from the complete example condition meeting the same SP as seen in the example (see Table 5). A t-test revealed significant differences between the performance ratings ($t(64) = -2.691$, $p = .009$) with higher values for those meeting the same SP as seen in the video-based modeling worked example compared to those who met another SP. Since not in all conditions data from learners having performed with the same SP is available, the SP cannot be included as an additional factor. Therefore, when the complete condition is included in an analysis, results on performance of the complex have to be interpreted with caution.

Table 5: Distribution of cases of students breaking bad news to the same or different SP previously seen in the worked example.

	complete (n = 15)	incomplete (n = 15)	incomplete + activity prompt (n = 14)	incomplete + explanation prompt (n = 8)	incomplete + activity prompt + explanation prompt (n = 14)
Different SP	7	14	14	8	13
Same SP	8	1	0	0	1

5.2 Results of the Research Questions

5.2.1 Effect of Video-Based Modeling Worked Examples on Complex Skill Acquisition

The first research question to answer is **RQ I: To what extent does the introduction of video-based worked examples affects the acquisition of a complex skill?**

To analyse the effect of the introduction of the video-based modeling worked example on performance of BBN, a one-factorial ANOVA with two factor steps (introductory text only vs. introductory text and video-based modeling worked example) and one dependent variable (complex skill of BBN) is performed.

There is a significant moderate effect of the introduction of a video-based modeling worked example on performance of BBN (F (1, 103) = 11.058, p = .001, η^2 = .097) with higher mean values of performance in the treatment group, see Figure 10.

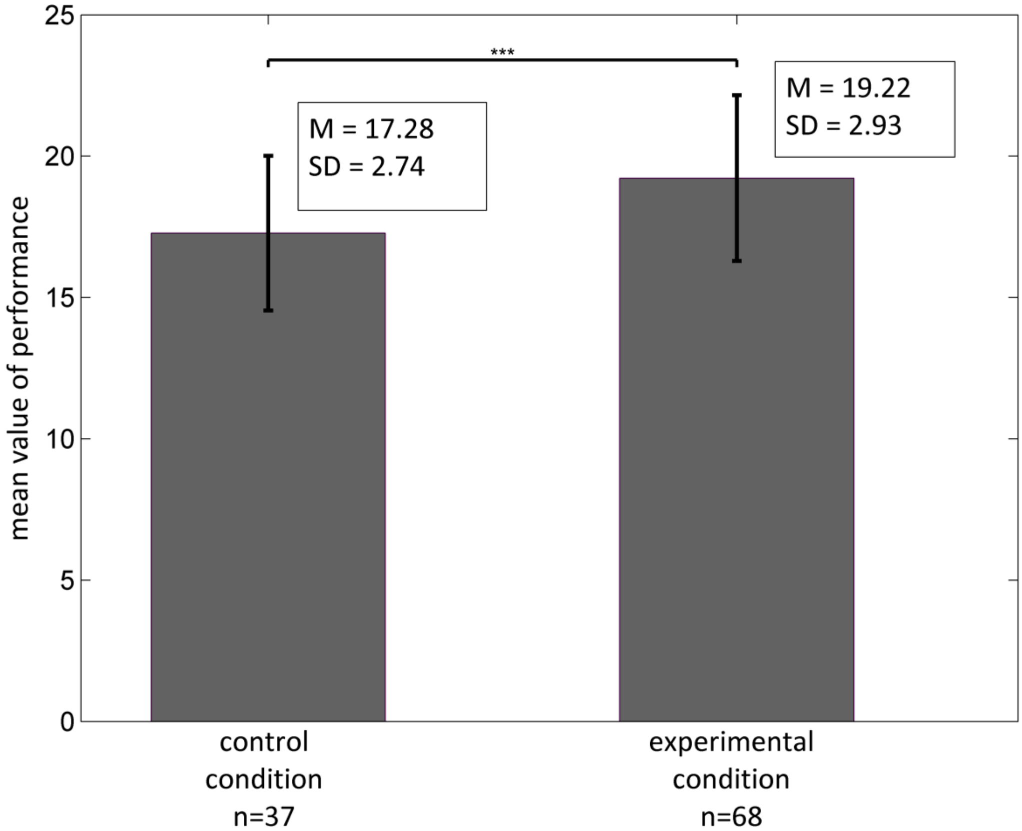

Figure 10: Mean usage of spikes steps with standard deviation during BBN in the control condition and in the experimental condition.

5.2.2 Effect of Self-Explanation Scaffolding on Process and Outcome of Video-Based Modeling Worked Examples

The second research question does no longer include the control condition from the semester before the intervention started, but focuses on the variation of completeness of the video-based worked example. **RQ II: To what extent does self-explanation scaffolding**

with incomplete video-based modeling worked examples affect processes and outcomes of video-based modeling worked examples?

As a first step, the complete example is compared against the incomplete example. To analyse the effect of completeness of the video-based modeling worked example on learning, a One-way ANOVA with all five factor steps of the modeling worked example and five dependent variables complex skill, knowledge acquisition (application-oriented), affect (post and gain), and cognitive load is performed (see Table 6). In a regression the post-values of the application-oriented knowledge test were predicted by the z-standardised pre-test values. The residuum is saved and used for all analyses to exclude prior knowledge.
The complete and incomplete condition is compared in its effects with the help of contrasts. Contrast one compares the complete condition with all incomplete conditions. In addition, the contrast between the complete worked example and the incomplete condition without further self-explanation prompts is compared to exclude effects of the different prompts (contrast two).

Table 6: Mean values and standard deviation of dependent variables.

		complete (*n* = 15)	incomplete (*n* = 15)	incomplete + activity prompt (*n* = 14)	incomplete + explanation prompt (*n* = 10)	incomplete + activity prompt + explanation prompt (*n* = 14)
complex skill	*Mean*	20.95	18.67	19.07	18.50	18.64
	SD	(2.220)	(3.177)	(2.841)	(3.100)	(2.977)
knowledge acquisition	*Mean*	.608	-.100	-.058	-.345	-.169
	SD	(.729)	(1.270)	(1.196)	(.547)	(.908)
affect (post)	*Mean*	14.46	13.36	12.92	14.30	14.21
	SD	(1.761)	(3.734)	(3.278)	(2.669)	(2.455)
affect (gain)	*Mean*	-.38	-.21	-.92	-.40	-.07
	SD	(1.193)	(1.528)	(1.038)	(2.547)	(1.207)
cognitive load	*Mean*	3.69	3.36	3.77	3.30	3.93
	SD	(1.251)	(1.216)	(1.235)	(1.767)	(1.269)

There are no effects of the condition when comparing all five instructional approaches in respect to learning and affect. But in line with the research question the complete and incomplete condition are compared by the contrasts defined above which show some significant effects of the experimental conditions. Those who worked with a complete video-based worked example performed the complex skill significantly better, contrasted to those with incomplete examples ($t(63) = -2.649$, $p = .010$), and also acquired more application-oriented knowledge ($t(52) = -2.258$, $p = .028$) (contrast 1, see Figure 11). Also the second contrast is significant in respect to learning by comparing the complete condition with the incomplete condition without further self-explanation prompts. In line with those results the complete condition has also a positive effect on the performance of the complex skill ($t(63) = -2.179$, $p = .033$). Application-oriented knowledge acquisition is not significantly different ($t(52) = -1.725$, $p = .091$). Neither of the contrasts reveals effects of the completeness of the video-based modeling worked example on negative affect or cognitive load.

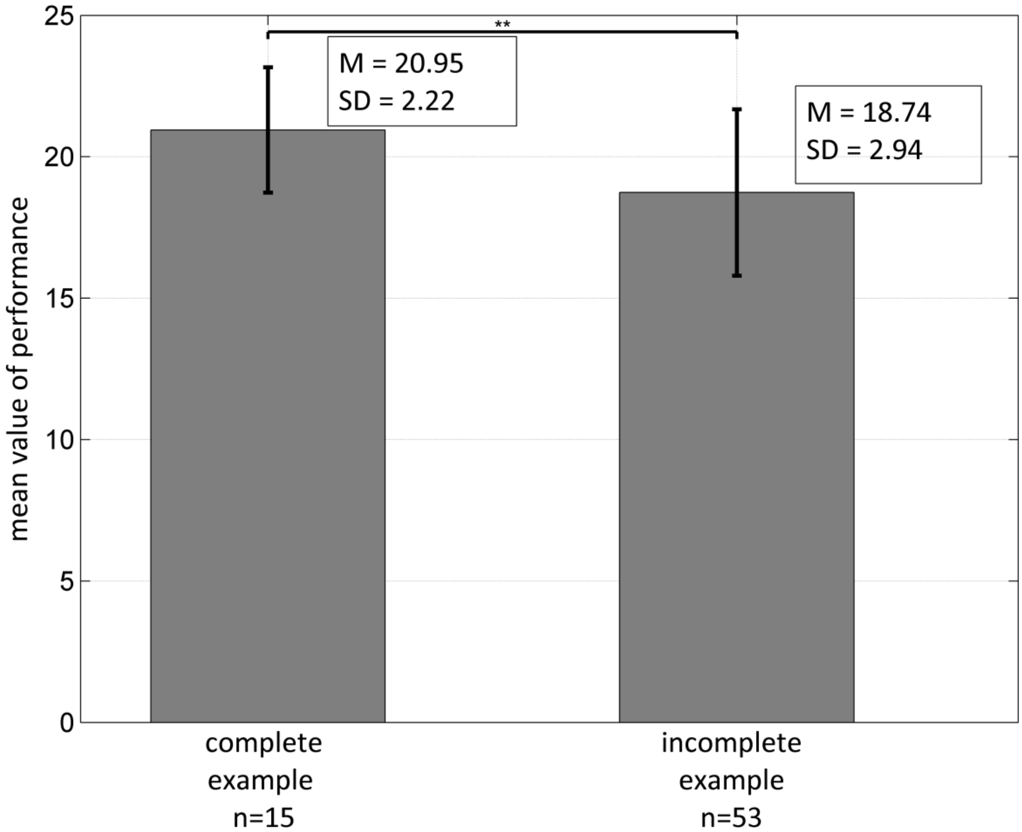

Figure 11: Performance with standard deviation between incomplete and complete video-based modeling worked examples.

By observing the values of the dependent variables it appears that there is a general effect of the video-based worked example on the reduction negative affect. To analyse this effect a One factorial ANOVA with repeated measurement with all five factor steps of the video-based modeling worked example and the repeated measurement of affect (inhibition and anxiety) is made. There is a significant moderate effect of the repeated measurement on negative affect ($F(1, 59) = 4.292$, $p = .043$, *partial* $\eta^2 = .068$) (see Figure 12).

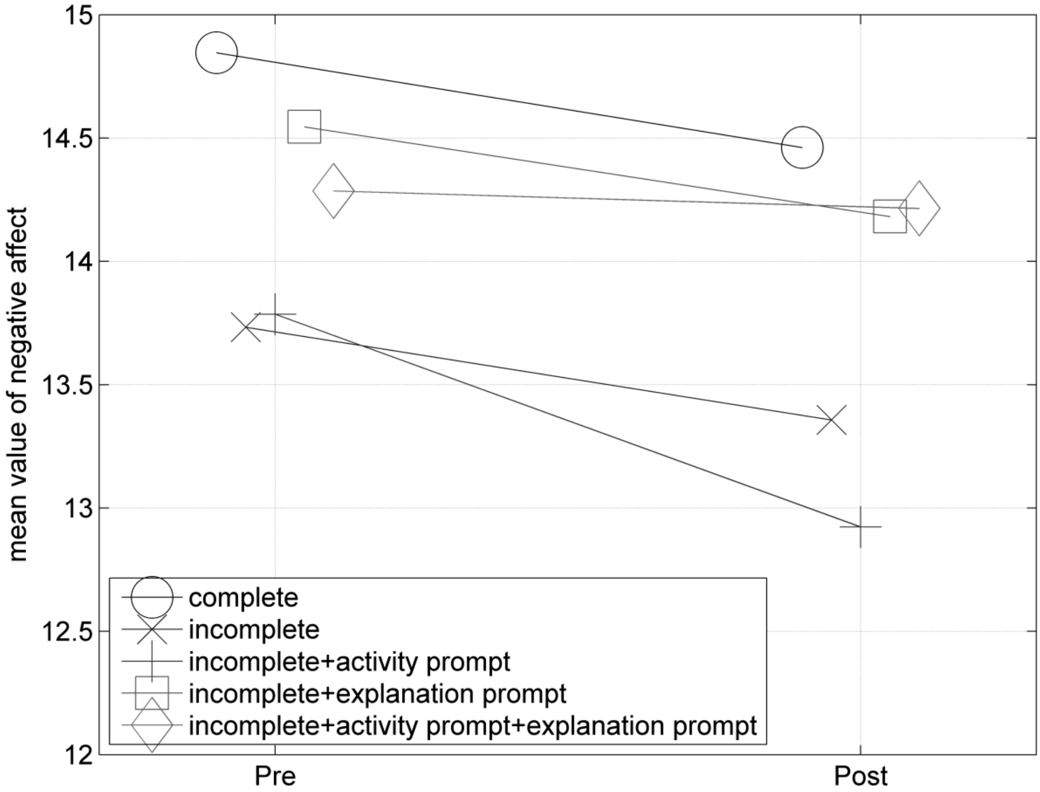

Figure 12: Negative affect between the treatment groups with repeated measurements.

5.2.3 Effect of Self-Explanation Prompts on Process and Outcome of Video-Based Modeling Worked Examples

As a next step, only the incomplete cases are analysed, comparing effects of different self-explanation prompts on learning. **RQ III: To what extent do self-explanation prompts during incomplete video-based modeling worked examples have an effect on learning process and outcome?**

To find out if there are differences in respect to learning in the incomplete condition due to different number and types of self-explanation prompts in the gaps, a MANOVA with the two factors activity prompt (yes vs. no) and explanation prompt (yes vs. no) and six dependent variables germane cognitive load, number of self-explanations (total and correct), complex skill, knowledge acquisition (application-oriented), and affect is performed (see Table 7 for the self-explanations; for the values of cognitive load, complex skill, application-oriented knowledge, and affect see Table 6).

There is neither a different main effect of the activity prompt (F (7, 36) = .217, p = .979; *Wilk's* Λ = 0.959, *partial* η^2 = .041) or the explanation prompt (F (7, 36) = 1.343, p = .259; *Wilk's* Λ = 0.7939, *partial* η^2 = .207), nor an interaction effect (F (7, 36) = .843, p = .559; *Wilk's* Λ = 0.859, *partial* η^2 = .141). There are no significant effects of the activity prompt in a video-based modeling worked example on either of the process or outcome variables. Explanation prompts in the gaps of incomplete examples show positive effects on the number of self-explanations (F (1, 45) = 5.949, p = .019, *partial* η^2 = .124) as well as on the number of correct self-explanations (F (1, 45) = 6.158, p = .017, *partial* η^2 = .128), see Figure 13. For the latter dependent variable the Levene test shows significant differences of variances (p = .001).

Therefore, the nonparametric Kruskal-Wallis-Test is calculated comparing the number of correct self-explanations and supports the finding by rejecting the null hypothesis (p = .040). No other significant effects of explanation prompts have been revealed.
The only significant interaction effect of the prompting conditions is found on the change in inhibition (F (1, 45) = 5.176, p = .028, *partial* η^2 = .110).

Table 7: Self-explanations during the incomplete video-based modeling worked example.

		incomplete (*n* = 15)	Incomplete + activity prompt (*n* = 14)	Incomplete + explanation prompt (*n* = 10)	Incomplete + activity prompt + explanation prompt (*n* = 14)
Self-explanations (SE)	Mean*	.33	.29	.90	.57
	SD	(.617)	(.469)	(.876)	(.646)
	Mean (correct SEs)*	.13	.14	.70	.36
	SD (correct SEs)	(.352)	(.363)	(.675)	(.633)

*Significant at the 5% level

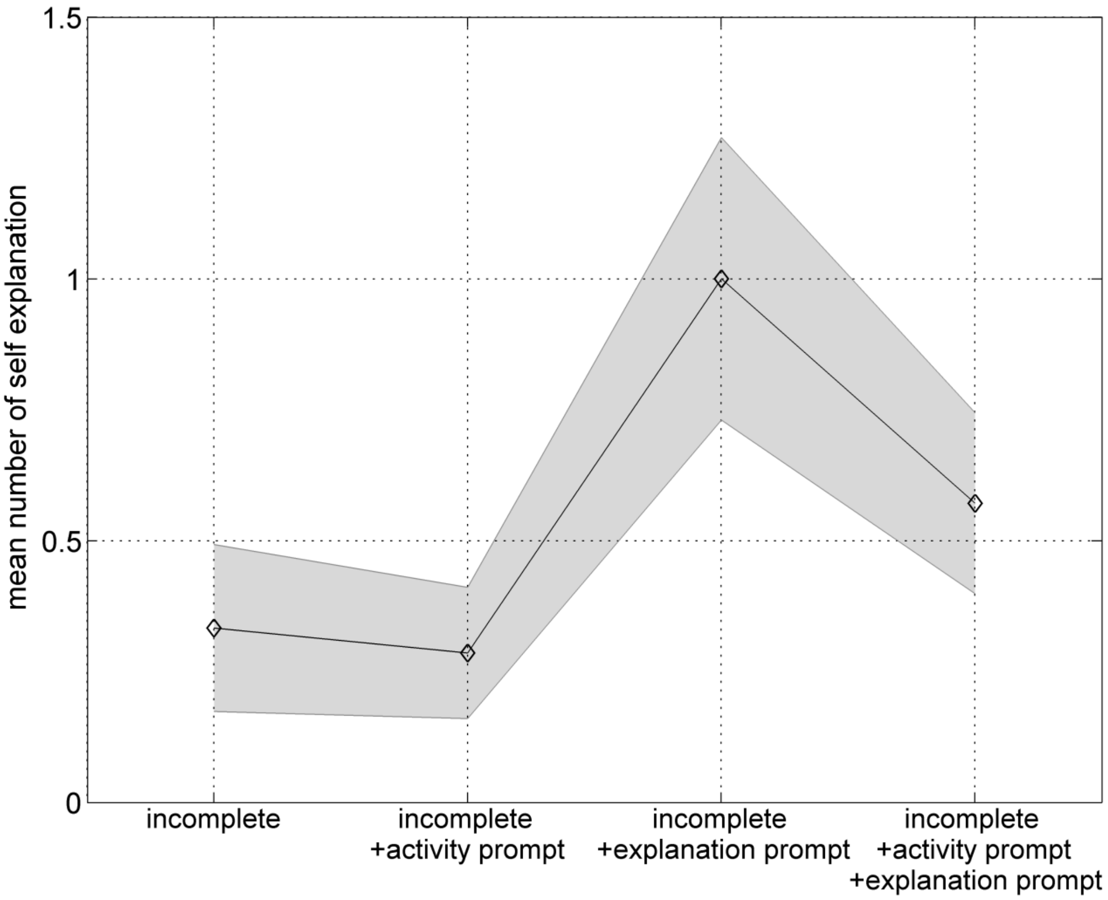

Figure 13: Mean number of self-explanation and standard error in the different prompting conditions.

5.2.4 Relation of Process and Outcome Variables of Video-Based Modeling Worked Examples

Finally, the relation of cognitive load and negative emotions and learning should be analysed. **RQ IV: To what extent are cognitive load and negative emotions related to the learning outcome when learning with a video-based modeling worked example?**

To answer this research question bivariate correlations have been performed to investigate relations of the process variables of learning cognitive load and number of (correct) self-explanations with the outcome variables performance and knowledge acquisition as well as negative emotions (anxiety and inhibition).

No significant correlations can be reported for either of the learning outcome measurements with the process variables. Cognitive load ratings show very weak correlations with application-oriented knowledge acquisition (N = 57, r = .022, p = .870) and negative correlations with performance ratings (N = 64, r = -.171, p = .178). The number of correct self-explanations correlates marginally with the performance ratings (N = 22, r = .068, p = .764) and even negatively with application-oriented knowledge acquisition (N = 19, r = -.278, p = .248). Negative emotions related to BBN and performance ratings show negative correlations prior to the worked example prior (N = 66, r = -.106, p = .396) and after (N = 64, r = -.088, p = .488) the video-based modeling worked example. The change in emotions related to BBN and performance is marginally correlated (N = 64, r = .027, p = .833). Application-oriented knowledge acquisition is marginally related to emotions prior (N = 57, r = .047, p = .726) and after (N = 57, r = .094, p = .486) the example as well as to change in emotions related to BBN (N = 57, r = .105, p = .436).

Also, there is no significant correlation of the number of correct self-explanations and cognitive load (N = 21, r = -.076, p = .745).

5.3 Summary and Conclusion on Results

The result do supports the hypothesis of research question I: There is a positive effect of the introduction of a video-based modeling worked example on the performance of the complex skills of BBN. This result shows that the training with a video-based modeling worked example has better effects on the skill of BBN than the schematic or theoretical approach of a text on the complex skill of BBN. Admittedly, there were more differences between the two groups besides the video-based modeling worked example. One major aspect was the time spent on task which was higher for the experimental condition in winter term 2009/10 compared to the control condition in the previous semester. On the one hand this aspect among other differences makes it impossible to deduce the effect to the video-based modeling examples only. On the other hand the significant moderate positive effect on the performance of BBN is a major argument for further investigations of the meachanisms of video-based modeling examples for complex skill acquisition. Therefore, the most effective variation of this instructional approach must be investigated in more detail in order to optimise learning effects.

One aspect that might add insights is the variation of completeness of the examples (RQ II). There is a general moderate reduction of inhibition and anxiety across all conditions. However, neither negative emotions nor cognitive load were affected by the completeness of the video-based modeling worked examples. The missing effect on cognitive load during the learning process is in line with expectations based on earlier research by Paas (1992). In respect to learning outcomes the participants in the complete condition acquired more application-oriented knowledge and performed the complex skill of BBN better than those of the incomplete conditions. The expectation based on empirical worked example research that an incomplete video-based modeling worked example should be more effective in respect to the learning process and outcome than a complete example can thus not be supported with the data

One explanation of the positive effect of the complete example on learning outcome is the skewed distribution of the SPs to conditions with those from the complete condition mostly meeting with the same person previously observed in the video-based modeling worked example while the learners from all other conditions in most cases met a SP that was unknown before. This confoundation is highly problematic and should be controlled in future studies. Another possible explanation for the contrary effect of completeness could be too little prior knowledge of the learners to handle the complex example. On the other hand the majority of 85% indicated that they had prior experience in the field of difficult communication situations and all students had already started the clinical part of their studies. Those prerequisites unfortunately prevent systematic analyses of effects of prior knowledge within this sample. Germane load ratings ranging between 3.3 and 3.9 [19](*SD* ranging from 1.2 to 1.8) also do not indicate major problems with the learning content.

After completeness did not show the expected effects on learning outcomes, the different prompting conditions within the incomplete video-based modeling worked example were compared in their effects on the acquisition of knowledge and skill as well as on the learning process (RQ III). There is an effect of the prompting conditions on the number of self-explanations as well as the number of correct self-explanations given by the learners with higher values for those learners who received explanation prompts. However, the mean output of self-explanations by the students is generally rather small. One reason might be that they had to type in their self-explanations instead of just pronouncing them as recommended by Roy and Chi (2005). Verbalisation is hardly possible though in a setting with several learners working in the same room as they would disturb each other which might have negative effects on learning. Also, the learning material in the study at hand differs from classical self-explanation research in that the number of details comprised in the video-based modelling worked example is much higher compared to text-based worked examples presenting a solution schema, although there are studies which successfully showed the self-explanation effect in multimedia settings and complex skill acquisition (Schworm & Renkl, 2007). However, the double-edged effect of prompts with even

[19] On the Likert scale applied a value of three translated to easy (einfach) and four translates to rather easy (eher einfach).

hindering effects of prompt on the elicitation of self-explanations has been reported before (Berthold et al., 2011) and therefore doesn't come as a total surprise.
As the mechanism of learning through self-explaining highly depends on the learner's prior mental model and its correctness (Kuhn & Katz, 2009), the explanatory power of the mere number of self-explanations is already limited. Besides this effect of explanation prompts on self-explanations, neither of the prompting conditions had a main effect on learning outcomes, negative emotions or cognitive load.

The reason why it is important that correct self-explanations are produced by the learners is their positive effect on learning and transfer (Shen & Tsai, 2009; Stark, 1999). Therefore, the relations of correct self-explanations given by the students during the learning process and learning outcomes were investigated (RQ IV). Bivariate correlations did not show the expected correlation between knowledge acquisition and performance of BBN with correct self-explanations during the learning process. Also, analyses of relations of learning outcomes with cognitive load and negative emotions (anxiety and inhibition) did not reveal any significant correlations. Generally, missing effects might be due to interfering unmeasured aspects that are crucial when learning with this special kind of learning material compared to classical worked examples. While this study on video-based modeling worked examples analysed variables and mechanisms reported in research on worked examples, there might be other concepts that focus more on the aspect of modeling and complex video-based examples. Explanations behind those missing effects of anxiety prior to the modeling worked example on the performance could thus point to a model experienced as too low in prestige. This aspect should be tested in a next step.

It can thus be summarised that the video-based modeling worked example had a positive effect on the performance of the complex skill of BBN and was also effective in reducing feelings of anxiety and inhibition linked to BBN. The students learned significantly better with the complete version of the worked example which however could also be partly due to performing with the same SP previously observed in the video-based worked example. Future studies should thus investigate effects of incomplete video-based modeling worked examples and control which SP the learners meet for BBN after the training. Within the incomplete examples, neither did self-explanation prompts have a main effect on learning nor did the number of (correctly) produced self-explanations. Generally, cognitive load, affect and learning outcomes did not show any connections.

Consequently, after the 1st study there are still many open questions to be answered on the mechanisms of video-based modeling worked examples. There are two lines of further approach for subsequent empirical investigations. Of the one part the experimental setup of this study must be improved with better control of confounding factors and adjusted measuring instruments. Kalyuga (2011) opts for a dual framework concentrating on intrinsic vs. extraneous load. Germane load is very difficult to separate from intrinsic load as it is conceptualised to describe the working memory ressources that are activated through intrinsic load. Leppink and colleagues (2014) tested a recently developed instrument by Leppink, Paas, Van der Vleuten, Van Gog, and Van Merriënboer (2013) containing scales for

all three load components in different learning contexts and with different instructional approaches. Only the scales for extraneous and intrinsic load could be confirmed by the data. Germane load as indicated to be assessed with the scale, did not correlate significantly with performance. Future studies should therefore follow the authors' recommendation and stick to measuring only extraneous and intrinsic load even though the relation of those is also not yet completely sorted out as they correlated in the cited study.
The role of cognitive load and its connection with negative emotions should furthermore be analysed with adjusted operationalization, adopting approved existing scales.

Of the other part, the focus should be placed on factors of research from the field of observational learning. Generally, missing effects of experimental variation may be due to the kind of model presented across all structural variations of the video-based worked example. The modeling performance had been selected from the video-taped BBN-performances as an average example of the control condition. Due to relatively little instructional support in the control condition, those performances are in general of rather low quality which is why the modeling worked example could be categorised as erroneous example showing a coping model. To understand the negative effect of incompleteness we have to take a closer look at the underlying mechanisms when learners work with video-based worked examples including a model and compare the erroneous example to a correct example. The effects of coping models embedded in video-based worked examples on learning are not yet determined empirically which is why the example structure should be tested again with different models, especially contrasting the coping model to a mastery model. Another aspect which has been discussed in research on observational learning is the prestige of the modeling agent which should be tested as well with a video-based modeling example.
In addition to affect related to the complex skill of BBN, more stable personal features of the learners like empathy should be tested for effects on learning with modeling examples. As different prompts did not show to have an effect on the production of self-explanations the mechanisms that support learning when working with video-based modeling worked examples are another open question to be investigated further, including measurements of the learner's relation to the model and the appraisal of it.

6 Research Questions of the 2nd Study

The aim of the 2nd study is to investigate the impact of different modelling cases and the personality trait empathy on the acquisition of complex skills with video-based worked examples, as well as on mediating factors like parasocial interaction of the learner with the model from the example. To add empirical results to the mechanisms behind modelling embedded in worked examples, modelling examples with different features in respect to expertise and relation to the observer are introduced and compared in their effects on learning, emotions and PSI. While research from the field of observational learning is inconclusive in respect to the optimal features of a model, typical models used in trainings in medical education so far are experienced doctors (Anolli et al., 2006). During this study different models will be compared in their effects on learning.

6.1 RQ I: Effect of Role-Model and Errors in Example on Learning Process with Video-Based Modeling Worked Examples

To what extent do errors in example and the role model presented affect PSI and the learning process when learning with video-based modeling worked examples?

One aim of the 2nd study is to further investigate the effect of different models in video-based worked examples and the correctness of their behaviour displayed on the learning process. What is the effect of watching a video-based worked example presenting either a BBN conversation of a coping model vs. mastery model in respect to the viewer's PSI with the doctor?

The intensity of PSI depends among other factors on the relation of the viewer and the persona and the authenticity of the latter (Klimmt et al., 2006). Classical social cognitive learning theory as well as recent research on observational learning underline the influence of the model as being perceived as sympathetic or not (van Ophuysen & Hannover, 2005). Besides the positive effects of a sympathetic model it is not clearly defined what the determinants of likability are besides negative effects of malevolence (Zillmann & Cantor, 1977). Furthermore PSI has been shown to be influenced by the degree to which the persona is perceived as realistic (Giles, 2002). Do the learners identify with a coping model and feel the pressure to perform better than him/her or do they on the contrary feel superior by separating from the model in the example? What is the effect of a student role model compared to a more prestigious professor role model?

The level of PSI is relevant because Klimmt and colleagues (2006) could show that high PSI resulted in more intense discussions following the viewing as well as more intense cognitions. Those cognitions might be categorised as self-explanations. Those have been shown to be positively influenced by learning with erroneous solutions (Siegler, 2002). It remains an open question if a coping model could also have a positive effect on self-explaining or if it might on the contrary impose more cognitive load on the learner and result in a reduced number of (correct) self-explanations.

In respect to the learning process, the aim of the study is to build on hints from worked example research on positive effects of erroneous examples and analyse its effect on cognitive load and the elicitation of self-explanations (Stark et al., 2011) with video-based examples showing a coping model.

Hypotheses: It is thus expected, that there is an effect of the persona from the modelling example with higher identification operationalized with higher levels of PSI with a student model as it bears more resemblance to the learners. It is furthermore expected that the video-based worked example with a student role model (that is more likely to the viewers) should evoke more self-explanations after the respective scenes of the example compared to a professor model.
It is an exploratory question whether there is an effect of coping model vs. mastery model on the number and correctness of self-explanations.
It is furthermore expected that learning with a coping student model example should lead to more extraneous load than a mastery student model presenting a correct solution.

6.2 RQ II: Effect of Role-Model and Errors in Example on Learning Outcomes of Video-Based Modeling Worked Examples

To what extent do errors in example and role model affect learning outcomes when learning with video-based modeling worked examples?

Worked examples can be seen as *expert models* per se (Schunk 1996 in (Crippen & Earl, 2007)). If the worked-example includes an example that is modeling, the question of the characteristics of the model presented arises. There are different findings in respect to the characteristics a model should have to be most efficient. Results from empirical studies range between best results for models of the same ability level (Schunk, 1996) to better effects of expert models (Baum & Gray, 1992; Boekhout et al., 2010). Social cognitive learning theory suggests that successful models that are similar to learners yield better results in respect to learning (Bandura, 1986).

As shown by Lim and colleagues (2009), relatively simple whole tasks are easily solved after learning with whole-task instructions. The worked-example can be categorised as whole-task. It can be supposed that the difficulty between the three examples varies to some extent, especially between the erroneous coping-case compared to the two expert cases that display correct examples. Differences in the outcome performance of the complex skill may thus also be due to differences in case difficulty induced by learning with a coping model compared to mastery models. It has been shown theoretically as well as experimentally that the assignment of incorrect worked examples can support learning (Stark et al., 2011). However, this positive effect does only occur if the learners' prior knowledge was high enough (Große & Renkl, 2007).

Hypotheses: It is thus expected that a video-based worked example with a mastery student model should have the best effects on learning results as the model is more likely to the learners compared to a professor model and at the same time displays the correct solution in contrast to a coping student model. In line with the expectation on performance of the complex skill it is expected that a correct example provided with a mastery model should have a positive effect on knowledge acquisition. The level of prior declarative knowledge should also have a positive effect on the learning outcomes as those learners who are novices in the field of BBN should profit more from the example with lower task difficulty (mastery model).

6.3 RQ III: Relation of Process and Outcome Variables of Role-Model and Errors in Video-Based Modeling Worked Examples

To what extent do affect, PSI, and cognitive load mediate the effects of the different video-based modeling worked-examples on performance?

6.3.1 RQ III.I: Effect of Role-Model and Errors in Video-Based Modeling Worked Examples on Affect as Learning Outcome

To what extent do time, errors in example, and role model affect affect when learning with video-based modeling worked examples?

It has been shown in the 1st study that the preoccupation with a video-based modeling worked example can reduce negative emotions linked to BBN (5.2.2 RQ II).
The 2nd study aims at a replication of this finding with enhanced scales for anxiety and inhibition and gain further insides in those effects by the comparison of different models displayed. As elaborated in 6.1 RQ I of the 2nd study, different role-models should provoke different levels of affective PSI in observing learners. It is expected that higher levels of affective PSI with the displayed models embedded in the video-based worked examples should lead to more cognitive activity after the observation (Klimmt et al., 2006). This in turn should have a positive effect on the handling of anxiety of BBN and inhibition to perform the complex skill.

Hypotheses: It is thus expected that working with a video-based modeling example should lead to a reduction of anxiety and inhibition related to BBN.
It is furthermore expected that higher levels of affective PSI during learning have a positive effect on the reduction of negative affect linked to BBN.

6.3.2 RQ III.II: Relation of Process and Outcome Variables when Learning with Video-Based Modeling Worked Examples

To what extent are personal traits, learning process variables, and learning outcomes of video-based modeling worked examples related?

After studying the effects of different role models and the effect of errors in a video-based modeling example on process (PSI, sympathy, cognitive load, self-explanations) and outcome variables (knowledge acquisition and performance of a complex skill), a subsequent aim of the 2nd study is to investigate the relation of those variables among themselves, including also the affective state before and after learning as well as the empathic disposition of the learners.

Empathy and PSI are closely connected concepts (Klimmt et al., 2006; Schramm & Wirth, 2010). While empathy is a relatively stable personality trait that is defined as the ability to identify and perceive emotions (Heberlein & Atkinson, 2009) PSI on the other hand is the state of being involved with a media persona on the cognitive and affective level in the process of observation. Especially involvement on the latter dimension is closely related to the concept of empathy by being defined by empathetic reactions, persona-generated own emotions, and mood contagion (Klimmt et al., 2006). PSI is furthermore conceptualised as involvement with the model displayed. This is closely linked to the acquisition processes of observational learning, defined by Bandura (1986) as *Attention* and *Retention* that are crucial for positive effects of modelling on learning outcomes. The appraised sympathy or likability of the model in the example has been found to have a positive effect on learning besides the level of PSI with the learner (Bandura, 1986; van Ophuysen & Hannover, 2005).

Another influencing factor of learning outcomes is cognitive load. Extraneous load has been shown to be negatively related to learning outcomes (Schwonke et al., 2011; Sweller, 1994). The same negative relation with learning outcomes has been reported for anxiety and motivation (Kim, 2012; Zeidner, 1998, 2007). On the contrary, self-explaning during the learning process has been found to have a positive effect on learning outcomes. This only holds true though if the exlanations by the learner are correct (Berthold & Renkl, 2009; Kuhn & Katz, 2009).

Hypotheses: It is thus expected that the PSI and empathy should correlate. With the concept of empathy as personality trait, the direction of influence should be empathic disposition influencing PSI and not the other way round.

It is furthermore expected that PSI as well as the perceived likability of the model from the example is positively correlated with learning outcomes.

In respect to extraneous load as well as anxiety and inhibition negative relations with learning outcomes are expected.

The learning process parameter (correct) self-explanations is expected to be positively related with knowledge acquisition and performance of the complex skill.

6.3.3 RQ III.III: Synthesis of Effects

Is there a model to sufficiently explain the effect of errors and role-model in the video-based modeling worked example on performance?

The aim of the 2nd study was to analyse if video-based modeling worked examples foster the acquisition of complex skills in general and more particularly how this process works. With the preceding research questions, specified aspects of the instructional approach, the learners' predisposition, learning process and learning outcomes have been investigated, presuming effects of the experimental variation in the video-based modeling worked examples on learning process and outcome.

Hypothesis: It is thus expected that more variance of the effect of errors in a video-based modeling worked example can be explained with the affective variables empathy, inhibition and anxiety prior to the example, and affective PSI compared to the effect of role-model in a video-based modeling worked example.
However, the basic learning mechanisms are expected to be the same for learning with all video-based modeling worked examples.

6.4 RQ IV Effect of Interventions on Complex Skill Acquisition

Did the intervention of the two studies succeed in improving the course?

The final aim is to compare the 1st and the 2nd study in order to derive a final conclusion on the improvement of learning by the interventions of winter semester 2009/10 and 2010/11.

Both interventions took place in the same setting with learners from the same population of medical students with only one year between them. To examine if the positive effect of the introduction of worked examples for complex skill acquisition can be enhanced by one of the experimental variations of the video-based modeling cases, it will be tested if the best condition from the 2nd study has the same or different positive effect on learning as the best condition from the 1st study (complete video-based modeling worked example with coping student model). It is expected that a mastery model leads to better learning results for the sample of relatively unexperienced medical students. Finally, in order to conclude on the total effect of the experimental intervention of including a video-based modeling worked example to foster complex skill acquisition, the control condition from summer term 2009 will be compared against the most effective condition of the 2nd study.

Hypothesis: It is therefore generally expected that both samples from the 1st and the 2nd study are comparable. It is furthermore expected that the same instructional approach should have the same effects on learning process and outcome during the 1st and the 2nd study. The most effective condition of video-based modeling example from the 2nd study is expected to have a relevant effect on the acquisition of the complex skill of BBN.

7 Method of the 2nd Study

The 2nd study took place during winter term 2010/11 in the same setting as the first data collection with the procedure remaining the same.

7.1 Participants

The 2nd study includes N = 100 participants that are medical students from LMU. Like the 1st study, the data collection was embedded in the *SP-Week* with students participated during their 4th year (71%). Two students still were in their 3rd year, and for 27 participants the information is missing. All data of the 2nd study were collected in winter term 2010/11. Like for the analyses of the 1st study, only those students were included that produced complete data sets. Incomplete data was on the one hand again caused by technical problems in respect to implementation of the learning material as well as data backup. On the other hand those data sets were excluded that were incomplete because the students did not voluntarily agree to be videotaped. Finally, those students with massive linguistic problems due to German not being their maternal language were excluded from the analyses. In the first weeks of the data collection the pilot phase led to some modifications in the learning material which is why only the data sets that were collected with the final version are included. Again, only the data of those learners the performance could be rated is included in the 2nd study. The mean age of the learners included is 23 years and seven month with a standard deviation of two years and eight month. Gender is relatively balanced within the whole dataset with 54% females. There is no significant difference between the video-cases of the factor errors in example (coping model vs. mastery model) in respect to the gender of the students that worked with them respectively (Pearson's Chi^2=.683, p=.409) and the SPs' gender the students met for the measurement of the complex skill of BBN (Pearson's Chi^2=.125, p=.723). Likewise, there is also no significant difference between the video-cases of the factor role-model (student vs. professor) in respect to the gender of the students that worked with them respectively (Pearson's Chi^2=.000, p=.991) and the SPs' gender the students met for BBN (Pearson's Chi^2=.076, p=.782). For the experimental design of the 2nd study, see Table 8.

Those who worked with the spikes-text while they were told not to are included in the data set because they are equally distributed to the different test conditions (Pearson's Chi^2=.221, p=.895) – even though they performed significantly worse (p=.005) with the SP than those who followed the instructions. The Students were, as in the 1st study, asked on a questionnaire prior to the study whether they already took part in Modul 2 and if they had delivered bad news before (*BBN before*). In contrast to the 1st study, they were explicitly asked to only indicate BBN conversations that took place outside of their studies. The item was changed to ‚I have already delivered bad news outside of Modul 2 for example in the scope of an internship.' (German item: 'Ich habe bereits eine schwierige Nachricht außerhalb von Modul 2 überbringen müssen, z.B. im Rahmen eines Praktikums.')[20]. 74% of the

---

[20]In the 1st study the item was *‚Ich habe bereits eine schwierige Nachricht übermittelt'*, which could also be answered with 'yes' by those students who delivered bad news in the scope of Modul 2.

participants indicated to already visiting Modul 2 before the training. For the other 26% the value is missing. The majority of 61% indicated that they had no prior experiences of delivering bad news outside of their studies. Only 13% already had delivered bad news to a patient. For 26% the value is missing. Previous knowledge[21] of the students is thus equally distributed to those who worked with the student video-cases coping model vs. mastery model respectively (Pearson's Chi^2=.288, p=.591) and those who worked with the mastery video-case student vs. professor respectively (Pearson's Chi^2=.137, p=.712), which is an important requirement when learning with worked examples. Prior experience is equally distributed among the test conditions (*Pearson's Chi^2* = .314, *p* = .855). Compared to the 1st semester, there are significantly less students in the 2nd study who indicated to have already delivered bad news (*Pearson's Chi^2* = 129.845, *p* = .000). The reason for this effect most probably is based in the reformulation of the question.

Table 8: Experimental design and dissemination of N = 100 cases to condition.

Factor	Video-based modeling worked example		
Role model	**Student model**		**Professor model**
Errors in example	**Coping model**	**Mastery model**	
Total number of cases: 100	n = 28 	n = 33 	n = 39 

7.2 Learning Environment and Procedure

The setting of this study is the same as described for the 1st study. The online learning environment the students worked with during the *SP-Woche* did only change in some details but mainly remained the same as in the 1st study. One difference is that the spikes-text had been translated into German while the students of the 1st study received the same text in English (for the complete workflow schedule of the 2nd study including all new aspects, see Table 9). The translation was made as a reaction of many requests by the students and to make sure there are no general problems in comprehension. The independent variation was changed according to the research questions of the second round of data collection and is implemented in phase 2 of the learning environment (7.3). While the main outcome variables as well as the coding of the self-explanations during the learning process remained the same or were only slightly changed, further survey instruments on affect, cognitive load,

[21]'I have already delivered bad news outside of Modul 2 for example in the scope of an internship.'

and PSI were added to phase 1 and 3 of the learning environment in order to gain more insights in personal learning factors (for more details, see 7.4).

Table 9: Time and workflow schedule with new aspects in the 2nd study with green background.

			time	content
Day one (Monday)			10 min	Short introduction in the project
			15 min	Text on the spikes in German
			2 min	Briefing for the simulation
	Online learning environment	Phase 1	5 min	Affect
			10 min	Application-oriented knowledge test (pre)
			10 min	Factual knowledge test (pre)
		Phase2	30 min	Video-based modeling worked example
				Items on cognitive load
		Phase3	7 min	PSI & affect
			10 min	Application-oriented knowledge test (post) + four items on affect + 1 item on cognitive load
			10 min	Factual knowledge test (post)
			1 min	Authenticity and Quality (validation of factors)
Day two (Tuesday)			20 min	Breaking bad news to a standardised patient

All learners from the 2nd study passed the whole process on day one and two. The experimental variation took place during phase 2 of the learning environment with the variation of three different video-based modeling worked examples.

7.2.1 Video-Based Modeling Worked Example

During this study, three different video-based worked examples are included and constitute the experimental variation. The example from the 1st study is included and will be labelled video A for better clarity. Video B, serving as a basis for the second video-based modeling example, is one of the modeling cases from the 1st study and was selected because the BBN performed is of high quality. The third example, referred to as video C, has been produced with a real physician who performs BBN with one of the SPs showing a high quality performance, as well. The structure and duration of the video-based modeling worked examples B and C are comparable with those of the example from the 1st study. While during the 1st study the structure of the learning material was varied and tested, the 2nd study relocated the focus of variation towards the content of the videos and standardised the structure. For all three video-based modeling worked examples an incomplete structure has been implemented, building on the incomplete condition of the 1st study, without adding further self-explanation prompts.

7.3 Experimental Conditions

In the 2nd study, three different videos have been used as a basis for the worked examples. During this study the participants worked with different video-based worked examples that were equal in respect to structure and duration[22] but differed in respect to the errors included in the performance of the complex skill (coping model vs. mastery model, see 7.3.1) and the role-model performing (student model vs. professor model, see 7.3.2). As studies on PSI showed significant effects of the model's gender (Klimmt et al., 2006) this factor is controlled by all-male models in the video-based modeling worked examples. With three different conditions, based on the respective video-cases A, B, C, the study plan has to be characterised as an incomplete 2x2 design. A fourth condition would have been composed on the basis of a coping professor role-model which would make sense methodologically but would be little convincing content wise. The latter aspect could lead to undesired reactions by the students not supporting and even hindering learning.

Figure 14: Screenshot of the learning environment (video-case B with the mastery student model) with an incomplete example.

In respect to the self-explanation prompts the kind of prompt given during the gap of the incomplete video-based modeling worked example was no longer varied during the 2nd study but the same for the whole group of medical students. During the gaps every participant was shown the question from the incomplete condition of the 1st study 'How should the doctor behave now?' (see Figure 14 showing the German formulation). The selection of scenes to be prompted in video B and C was made in a way to ensure high similarities in respect to content. For an overview of the structure of each of the videos in comparison, see Appendix 9: Overview of Experts' comments included in the three video-based worked examples.

[22] Duration of all video-case is between 10 and 12 minutes.

7.3.1 Errors in Example (Coping Model vs. Mastery Model)

The complex skill of BBN displayed in the examples is performed with different levels of expertise, operationalized as the number of correct steps spikes-steps performed by the models in the video-based worked-examples. Both models were selected to be similar in respect to obvious features: Both are male, of same age, and belong to the group of medical students. The two SPs displayed in the video-cases differ in respect to gender but are of similar age. Worked example B is based on one of the cases from the 1st study. It was selected because the performance of the complex skill was – in contrast to video-case A – one of the best during the semester. This is reflected by the number of spikes-steps that are performed by the doctor in the video: 24 steps have been coded which is far more than the mean number of spikes-steps performed in the 1st study (*mean* = 19.35, *SD* = 2.913) whereas the coping model in video-case A displays 16 out of 31 possible spikes-steps. For the exact numbers of steps shown in each of the examples respectively, see Table 10.

To check for validity of the variation of errors in the video-based modeling worked example one item was included on a seven-point Likert scale ranging from 1 = does not apply at all to 7 = totally applies: 'The conversation skills of the physician breaking the bad news were very good' (German item: ‚Die Qualität der Gesprächsführung, mit der der Arzt die schlechte Nachricht überbracht hat, war sehr hoch.'). Mean value for the coping student model was 3 (*SD* = 1.414) while the mastery student model was rated with a mean value of 6.11 (*SD* = 1.049). A t-test confirmed the validity of the factor errors in example with a large effect of the example's quality with higher quality allocated to the mastery student model compared to the coping student model ($t(63) = -9.883$, $p = .000$, $d = -2.656$).

Table 10: Number of coded spikes-steps of the three video-cases describing the quality of the complex skill displayed.

Role-model	**Student**		**Professor**	Maximum of steps per spikes
Errors in example	**Coping Model**	**Mastery Model**		
	A	**B**	**C**	
S	5	7	6	8
P	1	1	2	3
I	0	2	1	2
K	4	5	5	5
E	2	3	5	5
S	5	6	7	8
	17	**24**	**26**	31

7.3.2 Role Model (Student vs. Professor)

To experimentally vary the role-model without corrupting with effects of differences in the quality of the performance of the complex skill, two cases are contrasted that only differ in the prestige or role of the model displayed.

In our study the medical students were either shown a video with a medical student (video-case B) delivering the bad news to a SP or a professor (video-case C) (Table 10). The physician in the professor condition was familiar to the entire group of students so they could all recognise him as a real-life physician. Worked example C is based on a video-case that has been recorded especially for the study because the physician was supposed to be a real doctor and not a student from the 'SP Woche'. As in worked example B a high number of spikes-steps (26 steps) is performed.

To check for validity of variation of the role model in the video-based worked example one item was included to be rated on a seven-point Likert scale ranging from 1 = does not apply at all to 7 = totally applies: 'The person who took over the role of the physician performed very realistic' (German item: 'Die Person, die die Rolle des Arztes im Video übernommen hat, spielt diese Rolle sehr authentisch.'). Mean value of the mastery student model was 5.89 (*SD* = 1.133) while the mastery professor model was rated with a mean value of 6.34 (*SD* = 0.911). A t-test confirmed the validity of the factor role model, with a small to medium effect of the role model with higher values of authenticity allocated to the professor model compared to the student model ($t(84) = -2.029$, $p = .046$, $d = -.438$).

7.4 Dependent Variables and Data Sources

The variables for knowledge acquisition, complex skill, and (correct) self-explanations were again included in the test battery exactly as they were in the 1st study and will therefore not be described again. Only the discrepancies in respect to the dependent variables between the two study rounds will be reported. Those are new items for the measurement of affect and cognitive load, as well as completely new scales added to analyse empathy of the learners and their level of PSI with the model from the video-based worked example.

7.4.1 Knowledge Tests

Factual Knowledge

The test on factual knowledge acquisition is the same as applied during the 1st study with the identical four open questions and the identical order. The data collection of factual knowledge took once again place prior to the video-based modeling worked example in phase 1 of the learning environment and after it in phase 3 (see Table 9). The ratings of the data collection for factual knowledge range from a minimum of zero points to a maximum of 15 points reached for the pre-test ($M = 8.66$; $SD = 2.388$) and 14 points for the post-test ($M = 8.26$; $SD = 2.916$).). Like in the 1st study, all answers were coded by one to two trained coders who were trained to the coding scheme previously in order to reach similarity values with the predefined gold standard of every of the 15 categories of $\kappa > .6$ indicating substantial accordance.

Application-Oriented Knowledge

The video-based application-oriented knowledge test is the same as applied during the 1st study and reached minimum values of zero points and maximum values of eight points for working with video-case E ($M = 4.45$; $SD = 1.606$) and maximum nine points for working with

video-case Y (M = 4.24; SD = 1.91). Like in the 1st study, all answers were coded by one to two trained coders who were trained to the coding scheme previously in order to reach similarity values with the predefined gold standard of every of the 15 categories of $\kappa > .6$ indicating substantial accordance.
The order of the application-oriented knowledge test is equally distributed to the three different video-based worked examples (*Pearson's Chi²* = .832, p = .660).

7.4.2 Video

In the 2nd study seven SPs have been involved that have been randomly allocated to the students. Two male and five female SPs meet with students. Those were – except for two[23] - the same as in the 1st study (see Table 11). In contrast to the 1st study, there are neither difference between the conditions in respect to the students meeting the same or a different SP than they saw in the video-based worked example (Pearson's Chi²=028, p=.868). Furthermore, Mann-Whitney-U-Test for independent samples shows no significant differences between the SPs in respect to the performance ratings (p = .493).

Table 11: Standardised Patients involved in the data collection of the 2nd study.

SP			Video-Case			Number of BBN conversations analysed
Initials	Gender					
	male	female	A	B	C	
Be.*	x		0	4	1	5
Br.	x		12	5	9	26
Ho.**		x	1	3	2	6
S.		x	7	6	7	20
Hof.***		x	5	2	4	11
K.	x		2	9	10	21
W.		x	1	4	6	11
Total	3	4	28	33	39	100 (48 female/52male)

* Patient in video-case A (coping student model)
** Patient in video-case B (mastery student model)
*** Patient in video-case C (mastery professor model)

For the analysis, all available recordings of the BBN conversations were analysed with videograph® by time-sampling with the same coding scheme implemented during the 1st study. Duration of the performances ranged between 4,6 and 18 minutes per student. Analogically to the 1st round of data collection, all performance-recordings are coded by two to four trained coders with a minimum inter-rater reliability of $\kappa > .6$ for the different categories. The mean time spent for BBN with a SP by the learners is 10 minutes with a standard deviation of 2 minutes 20 seconds. For the mean performance ratings and time spent on BBN per factor group, see Table 12.

[23] The new SPs are one male (Mr. K.) and one female (Mrs. W.).

Table 12: Mean ratings and duration of performance in the different conditions of video-based modeling worked examples (2nd study).

		A - coping student model (N=28)	B - mastery student model (N=33)	C - mastery professor model (N=39)
Performance	Mean	19.39	22.82	20.38
	SD	(3.304)	(3.087)	(3.023)
Duration of BBN in minutes	Mean	9.348	10.080	10,389
	SD	(2.454)	(2.216)	(2.498)

7.4.3 Questionnaire

Affect

Anxiety is the emotion that has been most often assessed in the scope of learning. In the 1st study only one item was included in the questionnaires explicitly asking for anxiety in respect to BBN (see 4.4.4). To verify reliability a new scale for anxiety has been composed for the 2nd study with items from different sources that applies to the given scenario. Ten items are based on the Test Emotions Questionnaire TEQ (Pekrun, Götz, Perry, Kramer, & Hochstadt, 2004) and one item is taken from the 1st study. One exemplary item added to the scale is ‚I am so nervous that I would rather delegate the delivery of the bad news to a colleague.‘. Like in the 1st study, all items had to be rated on a five-point Likert scale from 1: *'Do not agree'* to 5: *'Agree'*. Sufficient reliability of the scale is given with Cronbach's alpha ranging between α = .863 (pre) and α = .913 (post) for the two measuring times. For all German items of the anxiety scale, see Appendix 10: Items of the anxiety scale applied in the 2nd study. The mean value of the complete scale for the whole dataset (including those with performance rating) is 30.73 (*SD* = 6.87; *min* = 12; *max* = 52) for anxiety to break bad news prior to the modeling worked example and 28.42 (*SD* = 7.94; *min* = 13; *max* = 51) for anxiety of the complex skill after the example.

Some changes in the scale for inhibition were made, including four additional items and a reformulation of two out of the three old items in order to shape the concept of inhibition and make explicit links to the BBN situation (see Appendix 11: Inhibition scale used in the 2nd study in comparison to scale used in the 1st study.). One exemplary item added to the scale is ‚I try to keep difficult conversations with patients as brief as possible.‘. Two items (1 and 4) are negatively formulated. Sufficient reliability of the scale is given with Cronbach's alpha ranging between α = .739 (pre) and α = .767 (post) for the two measuring times (only those cases included with a performance rating). The mean value of the complete scale for the whole dataset (including those with performance rating) is 24.19 (*SD* = 4.76; *min* = 13; *max* = 34) for inhibition to break bad news prior to the modeling worked example and 23.79 (*SD* = 4.77; *min* = 13; *max* = 33) for inhibition after the example. To compare the two studies in respect to the rating of the items that appeared in both questionnaires a one-way ANOVA with the factor study (1st study vs. 2nd study) and the matching item for anxiety (pre and post) and three matching items for inhibition (pre and post, respectively) as dependent variables showed no significant differences between the items of the inhibition scales of both semesters.

Empathy

The concept of empathy is one of the measurements added to the test battery of the 2nd study. It has been measured by self-assessment. The items of the self-assessment scale target the general level of empathy across different situations. Prior to the course that served as a framework for the data collection, all students visited another seminar where they were asked to fill in a questionnaire that asked for their empathy. As empathy is a trait that should not vary too much between different situations, the removal from the course was a good possibility to reduce the number of questionnaire items during the learning environment.

Empathy was measured with the E-Skala (Leibetseder et al., 2001) with 25 Items on a five-point Likert scale (α = .902) ranging from 1=does not apply to 5=applies. Both subscales are confirmed by sufficient reliability scores: 1. *Readiness for Empathy*[24] (13 items, two negatively formulated, α = .899): Ability and willingness to put oneself in the position of a fictional other in respect to experience and behaviour. An exemplary item for this scale is ‚When I'm reading an interesting story I imagine how I would cope with such a situation' 2. *Social Concern*[25] (12 items, one negatively formulated, α = .831): Emphatic behaviour in real situations. An exemplary, negatively formulated item for the second scale is 'Other peoples' mishaps typically do not touch me'. For all German items of the questionnaire see Appendix 12: E-Skala for self-assessment of empathy by Leibetseder and colleagues (2001). Mean value of the scale *Readiness for empathy* for the whole dataset (including those with performance rating) is 41.99 (*SD* = 11.04; *min* = 13; *max* = 63) which translates in a mean value per item of 3.23. The mean value of the scale *Social Concern* for the whole dataset (including those with performance rating) is 39.54 (*SD* = 7.83; *min* = 12; *max* = 57) which translates in a mean value per item of 3.3.

Parasocial Interaction

Another newly added scale is based on the Process Scales by Schramm and Hartmann (2008) to measure parasocial interaction (PSI) as an extension of *normal* social interaction (Giles, 2002). PSI is measured to investigate the effect of the training session on the perceived relation of the doctor to the observed model (doctor) and in which way the learner is affected by the observation.

It can be differentiated between high vs. low PSI. The level or intensity of PSI can moderate learning outcomes via its effect on the intensity of post-viewing discussions and cognitions (Klimmt et al., 2006). In order to enhance the informative value by adopting higher order analyses, it will not be differentiated between high and low PSI by reducing the level of measurement to nominal scales. Instead, the interval level of measurement is kept so that analyses of variance are facilitated.

---

[24] Translated ftom the German label ‚Einfühlungsbereitschaft'

[25] Translated from the German label ‚Betroffenheit'

To make research on PSI comparable Schramm and Hartmann (2008) developed the PSI Process Scales. PSI can thus be measured on three levels that incorporate sub-dimensions: a) Cognitive level with up to six dimension, b) Affective level with up to five dimensions, and c) Behavioural level with up to three dimensions. For each subscale up to eight items are provided which makes a total of 112 items on the scale to be rated on a seven-point Likert scale ranging from 1 = I do not agree at all to 7 = I totally agree. The questionnaire can be applied for all interactions across different contexts in the complete version or in parts.

The short version applied in this study includes 13 Items that have been selected on the basis of plausibility (items that seem to fit with the stimulus material) and diversity (items that represent different aspects/facets of the construct). Only cognitive and affective scales have been included: Seven items on Affective PSI with four negatively scaled items, including the exemplary item 'What the doctor said or did didn't provoke any emotions.' and six items on cognitive PSI including one negatively scaled item 'I never actually reflected on what the doctor should do or say next.'. For all German items included in the test battery see Appendix 13: PSI scale based on Process Scales by Hartmann & Schramm (2008). The items have been reformulated to match the special situation and the persona of the video-cases and target the PSI of the viewer with the doctor displayed in the video and not with the patient. For the statistical analyses, the total item scores have been divided by the number of items per scale, respectively, resulting in comparable values between one and seven points.

All participants indicated their answers on a computer-based questionnaire after they worked with the video-based modeling worked example. As every student only watched one out of the three video-based modeling worked examples their level of affective and cognitive PSI is directly linked to the doctor in the example. Both subscales are reliable with α = .659 for the scale on cognitive PSI and α = .789 for the scale on affective PSI. Mean values for the whole dataset (including those with performance rating) are with 5.41 (SD = 0.81; min = 3.67; max = 7) higher for cognitive PSI compared to 4.32 (SD = 1.04; min = 1.57; max = 7) for affective PSI, indicating higher levels of agreement on the cognitive level.

One additional item on sympathy for the observed model was included in order to filter the data because it was not possible to operationalize the affective processes of PSI free from valence. This is necessary as, in contrast to original work on PSI where the participants were – following the rating of likability – only given scales of sympathy and empathy or antipathy and counterempathy, in our questionnaire only items on a positive perception were included. Therefore, the students additionally asked to indicate how likable they perceived the doctor during the video: 'Please, indicate how likable you perceived the doctor during the video'. The item was to be rated on a six-point Likert scale ranging from 1 = very dislikable to 6 = very likable. Mean value for the whole dataset is 4.63 (SD = 1.3; min = 1; max = 6).

Cognitive Load

During the 2nd study the scale to assess cognitive load was exchanged. Instead of the two items targeted to rate the difficulty of the text and the video-based modeling worked example in respect to germane load, cognitive load was now assessed with four items at two measuring points during the video-based modeling worked example to enhance validity: One item for intrinsic load and three items for extraneous load. At both measuring points the items were identical and based on the assessment scale by Maria Opfermann (2008) (for the German items see Appendix 14: Items to measure cognitive load based on the scale by Opfermann (2008).). All items were rated on a nine-point Likert scale ranging from 1 (extremely easy) to 9 (extremely difficult). The item on intrinsic load is 'How easy or difficult do you consider breaking bad news at this moment?' while an exemplary item for extrinsic load is 'How easy or difficult is it for you to work with the learning environment (the videos, the expert's comments, …)?'. For analyses, mean values of all items were used to avoid corrupting effects of the number of items per scale (accumulated values of one scale divided by number of items on the scale). This results in one value per scale with a possible range of values between 1 and 9. Mean values for the whole dataset are 4.03 (SD = 1.37; min = 1; max = 9) for extraneous load and 6.03 (SD = 1.57; min = 2; max = 9) for intrinsic load. Reliability is sufficient for both scales, with α = .877 for extraneous load and α = .849 for intrinsic load, for both measuring times together and α = .705 (for measuring time one) / α = .879 (for measuring time two) for both measuring times of extraneous load.

7.5 Statistical Analyses

For the 2nd study the same statistical parameters as in the 1st study apply for the evaluation of effect sizes as well as the approach of exclusively including only the data sets that comprise a performance rating.

In addition to the statistical analyses run in the 1st study regressions have been performed.

Within a regression R^2 can take values between 0 and 1 with values from 0.0196 indicating small effect sizes, values starting from 0.1300 indicating moderate effects and values starting at 0.2600 indicating large Effects. The standardized partial coefficients (*Beta*) describe the comparative relevance of the predictor. The larger the beta values, the more variance can be explained by the associated variable.

8 Results of the 2nd Study

Before targeting the actual research questions, preliminary analyses of comparability between the factor steps as well as between the two rounds of the study will be reported.

8.1 Preliminary Results

In contrast to the 1st study, bivariate correlation of duration of BBN and performance showed medium to strong significant correlations (N=100, r=.437, p<.000), indicating better ratings of the complex skill of BBN connected to longer duration of the conversations with the SP. Even though the mean duration of BBN in the 2nd study was a bit longer than during the 1st study, a t-test showed no significant effect of the semester on duration ($t(162) = -1.696$, $p = .092$, $d = -.271$). However, the dissemination of duration of BBN is equal between the different video-based modeling worked examples ($F(2, 97) = 1.568$, $p=.214$).

Due to time constraints on the side of the SPs, technical problems with the video recording and missing data from some participants, the distribution of SPs to test condition is not totally balanced for the 2nd study. There is a significant difference between the coping model and mastery model condition in respect to the SP that followed on the next day (Pearson's $Chi^2=15.192$, p=.019). Therefore, the SP will be included as a factor in the analyses of effects of the video-based modeling worked example.

There are also different distributions in respect to visiting Modul 2 before with significant less student in the 2nd study indicating to not yet having participated (*Pearson's* $Chi^2 = 10.584$, $p = .001$).

The declarative knowledge test has again been shown to be problematic in the same sense as during the previous data collection with lower mean value of the post-test compared to the mean pre-test values. Therefore only the pre-test value is included in the analyses.

In contrast to the 1st study, a t-test showed a small but not significant effect ($t(82) = 1.765$, $p = .081$, $d = 0.39$) of the order of the video-cases within the application-oriented knowledge test with the order case E – case Y resulting in higher means of knowledge acquisition. The effect size gives a hint to a relevant influencing factor which is why, in contrast to the 1st study, the factor order of the videos of the knowledge test will be included in all analyses made with the data from the knowledge test to avoid confoundation of it.

Table 13: Affect ratings prior to the video-based modeling worked example.

		coping student model (n = 23 / 7)	mastery student model (n = 29 / 21)	mastery professor model (n = 37 / 26)
Anxiety pre	Mean	30.26	31.03	30.78
	SD	(5.79)	(7.66)	(7.00)
Inhibition pre	Mean	22.86	23.67	24.96
	SD	(3.44)	(5.36)	(4.57)

There are no significant differences between the factor groups in respect to ratings of anxiety ($F(2;86) = 0.082$, $p = .922$) and inhibition ($F(2;51) = 0.735$, $p = .485$) prior to the video-based worked example (see Table 13). Contrasts comparing the coping student model against the mastery student model and the mastery student model against the mastery professor model show to be not significant for anxiety as well as for inhibition.

8.2 Results of the Research Questions

8.2.1 Effect of Role-Model and Errors in Example on Learning Process with Video-Based Modeling Worked Examples

The first research question of the 2nd study to answer is **RQ I: To what extent do errors in example and role model affect PSI and the learning process when learning with video-based modeling worked examples?**

The first part of the question is investigated first by concentrating on the effect of errors in example (coping student model vs. mastery student model) and role model (mastery student model vs. mastery professor model) on PSI (cognitive PSI and affective PSI) and likability of the model (sympathy). For the mean values and standard deviations, see Table 14.

Table 14: PSI- and sympathy-ratings between the video-based modeling worked examples.

		coping student model (n = 23)	mastery student model (n = 26)	mastery professor model (n = 34)
cognitive PSI *	Mean	5.739	5.43	5.17
	SD	(.685)	(.866)	(.787)
affective PSI *	Mean	4.282	4.660	3.992
	SD	(.871)	(1.123)	(.972)
sympathy **	Mean	3.22	5.27	5.09
	SD	(.850)	(.919)	(1.055)

* Significant at the 5% level

** Significant at the 1% level

To analyse the effect of the three different video-based worked examples on PSI and likability, a MANOVA with three factor steps of different worked examples (coping student model vs. mastery student model vs. mastery professor model) and three dependent variables (cognitive PSI, affective PSI and sympathy) is performed, including only those datasets with performance rating.

There is a statistically significant large effect on the variation in video-based modeling worked example on PSI and sympathy (F (6, 156) = 11.560, p < .000; *Wilk's* Λ = 0.479, η^2 = .308). The worked example has a statistically significant medium effect on cognitive PSI (F (2, 80) = 3.617; p = .031; *partial* η^2 = .083), Affective PSI (F (2, 80) = 3.320; p = .041; *partial* η^2 = .077), and a large effect on sympathy (F (2, 80) = 34.520; p < .000; *partial* η^2 = .463), see Figure 15.

To find out if the difference in the role model (student vs. professor) or the difference in errors in example (coping vs. expert) can explain the effects on PSI and sympathy for the model, two more MANOVAs are calculated. The MANOVA with factor role model in video-based modeling worked example (student vs. professor) shows the differences in affective PSI with a medium effect (F (1, 58) = 6.093; p = .017; *partial* η^2 = .095). No effect of the factor role model on cognitive PSI (F (1, 58) = 1.480; p = .229; *partial* η^2 = .025) or sympathy (F (1,

58) = .484; p = .489; *partial* η^2 = .008) can be reported. The MANOVA with factor errors in video-based modeling worked example (coping vs. expert) shows no differences in affective PSI (F (1, 48) = 1.707; p = .198; *partial* η^2 = .035) and cognitive PSI (F (1, 48) = 1.893; p = .175; *partial* η^2 = .039) but a large significant effect on sympathy (F (1, 48) = 65.216; p < .000; *partial* η^2 = .581) with higher levels of likability reported of the expert student model compared to the coping student model.

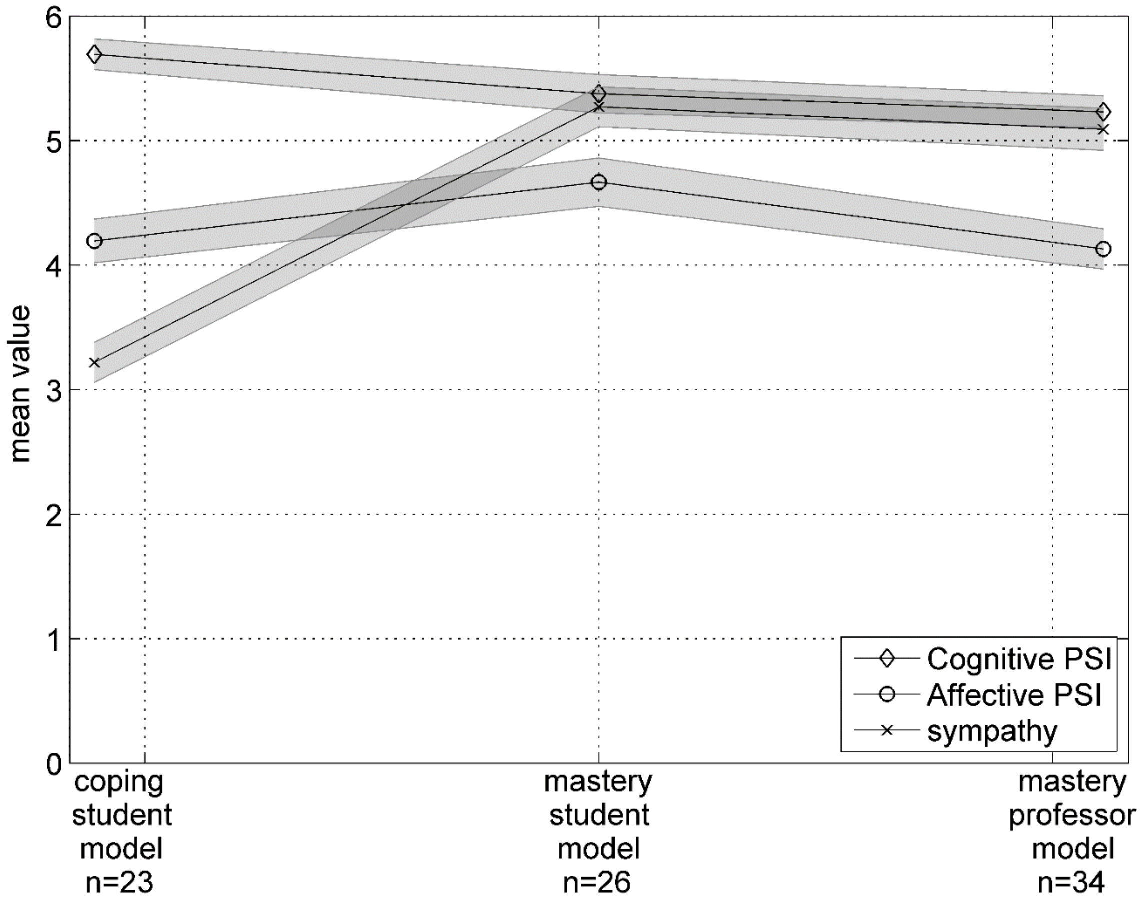

Figure 15: PSI and sympathy ratings between the three different video-based worked examples (with standard error).

As a second step the effect of the factor errors in video-based modeling worked example (coping student model vs. mastery student model) on cognitive load (extraneous load and intrinsic load) is analysed. The effect of the factor role model is not investigated because no effect is expected on cognitive load.

Table 15: Mean values of extraneous and intrinsic load of the student model worked-examples.

		coping student model (n = 28)	mastery student model (n = 30)
extraneous load	Mean	4.05	3.93
	SD	(1.53)	(1.17)
intrinsic load	Mean	6.07	6.18
	SD	(1.39)	(1.94)

To analyse the effect of errors included in the video-based worked examples on cognitive load, a One-way ANOVA with two factor steps of errors in the modeling worked examples (coping student model vs. mastery student model) and two dependent variables (extraneous load and intrinsic load) is performed, including only those datasets with performance rating.

There is no effect of errors during the video-based modeling worked example on extraneous load (F (1, 56) = 0.114; p = .737; η^2 = .002). Also no significant effect of the errors in the example on intrinsic load (F (1, 56) = 0.063; p = .803; η^2 = .001). Generally, the values range around medium levels of cognitive load (minimum of the scales would be 1 and maximum 9, respectively, see Table 15).

The third part of RQ I of the 2nd study targets the production of (correct) self-explanations by the learner while working with a different video-based modeling worked example.

Table 16: Content of gaps in incomplete worked examples.

		coping student model (n = 28)	mastery student model (n = 33)	mastery professor model (n = 39)
self-explanations	Mean	.96	.58	.64
	SD	(.999)	(.830)	(.778)
correct self-explanations	Mean	.68	.45	.33
	SD	(.772)	(.564)	(.621)
Knowledge during gaps	Mean	6.43	5.82	6.36
	SD	(1.709)	(2.417)	(2.096)

To analyse the effect of the three different video-based worked examples on self-explanations, a MANOVA with three factor steps of different worked examples (coping student model vs. mastery student model vs. mastery professor model) and three dependent variables (number of (correct and total) self-explanations and knowledge during gaps) is performed. Only those data sets are selected that include a performance rating (Table 16).

There is an effect of the different video-based worked examples on the number of (correct) self-explanations, and knowledge during the gaps that is not statistically significant (F (6, 192) = 1.717, p = .119; *Wilk's* Λ = 0.901, *partial* η^2 = .051). Between subjects effects are not significant for none of the dependent variables but indicating small effect sizes for self-explanations (F (2, 97) = 1.745; p = .180; η^2 = .035) and correct self-explanations (F (2, 97) = 2.314; p = .104; η^2 = .046). As the Levene test is significant for the comparison of correct self-explanations, two nonparametric comparisons are made (Mann-Whitney-U-Test) which are non-significant for the factor role model (p = .223) as well as for the factor errors in example (p = .306).

The errors in example and role model conditions are compared in their effects with the help of contrasts. Contrast one compares the coping student model example against the mastery student model example. The second contrast compares both mastery models that differ in their role model characteristics.

None of the contrasts is significant, only the small effect of contrast 1 ($t(97) = 1.755$, $p = .082$, $d = .311$) would be significant in respect to the number of self-explanations produced if tested one-tailed with more self-explanations by those who worked with the erroneous coping student model case.

8.2.2 Effect of Role-Model and Errors in Example on Learning Outcomes of Video-Based Modeling Worked Examples

The second research question of the 2nd study to answer is **RQ II: To what extent do errors in example and role model affect learning outcomes when learning with video-based modeling worked examples?**

As a first step, the effect of the video-based modeling worked examples (coping student model vs. mastery student model vs. mastery professor model) on the complex skill of BBN (performance) is tested.

Table 17: Performance ratings of learners working with different video-based worked examples.

		coping student model (n = 28)	mastery student model (n = 33)	mastery professor model (n = 39)
Performance **	Mean	19.39	22.82	20.38
	SD	(3.304)	(3.087)	(3.023)

** Significant at the 1% level

To analyse the effect of the three different video-based worked examples on performance of the complex skill of BBN, an one-way ANOVA with three factor steps of different worked examples (coping student model vs. mastery student model vs. mastery professor model) and one dependent variable (performance of the complex skill) is performed.
There is a significant large effect of the video-based modeling worked example on performance of the complex skill of BBN (F (2, 97) = 10.008, p = .000, η^2 = .171), see Figure 16.

To rule out any effects of the SP the learners performed with, an one-way ANOVA with SP as an additional factor to the video-based modeling worked example was made, showing that there is no significant effect of the SP the students performed the BBN conversation with on the quality of their performance ($F(6, 80) = .943$, p=.469, *partial* η^2 = .066) in addition to the large significant effect of the treatment factor video-case (F (2, 80) = 7.244, p = .001, *partial* η^2 = .153). Also, the (large) interaction effect of video-case in the learning unit and SP during BBN conversation on competence is not significant ($F(11, 80) = 1.281$, p=.250, *partial* η^2 = .150). Therefore the SP will not be taken into account in further analyses.

To examine to what extend the factor role model (student vs. professor) or the factor errors in example (coping model vs. mastery model) can explain the large effect, two t-tests are calculated.

The video-case factor errors in example (coping model vs. mastery model) has a large statistically significant effect on performance ($t(59) = -4.182$, $p = .000$, $d = 1.07$) with more spikes-steps performed by those students who worked with the video-based worked example showing the expert model's performance. The video-case factor role model (student vs. professor) also has a large statistically significant effect on performance ($t(70) = 3.371$, $p = .001$, $d = 0.80$) with more spikes-steps performed by those students who worked with the video-based worked example with the student role model. According to this, the effect can be explained by the two video factors errors in example and role model in nearly equal measure.

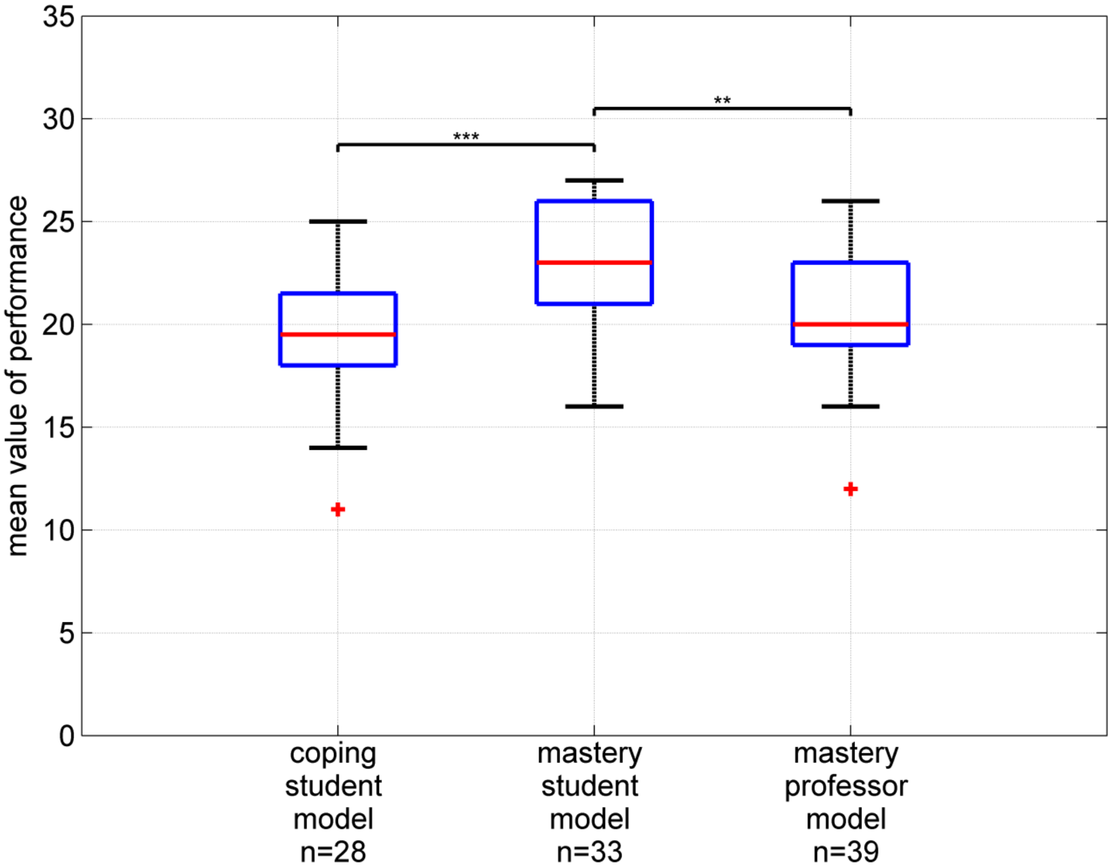

Figure 16: Performance between students watching the three video-based modeling worked examples.

As a next step, the effect of the video-based worked examples, more precisely the factors errors in example and role model (coping student model vs. mastery student model vs. mastery professor model) on knowledge acquisition by involving prior declarative knowledge is tested.

To analyse the effect of the three different video-based worked examples on knowledge acquisition of BBN, a two-factor ANCOVA with three factor steps of different worked examples (coping student model vs. mastery student model vs. mastery professor model), two factor steps of the order of videos within the knowledge test (Y-E vs. E-Y), prior declarative knowledge as covariate, and one dependent variable (application-oriented knowledge acquisition[26]) is performed.

Table 18: Knowledge acquisition between the video-based modeling worked examples, differentiating the order of the knowledge test video-cases.

		coping student model (*n* = 24)		mastery student model (*n* = 27)		mastery professor model (*n* = 33)	
		E-Y	Y-E	E-Y	Y-E	E-Y	Y-E
Knowledge acquisition *	Mean	.732	.056	-.085	-.200	.031	-.385
			.366		-.149		-.144
	SD	(.965)	(.710)	(.902)	(1.105)	(1.042)	(.712)
			(.886)		(1.003)		(.927)

*Significant at the 5% level

There is a moderate effect of the different video-based modeling worked examples on knowledge acquisition (F (2, 77) = 2.914, p = .060, *partial* η^2 = .070) that is only significant if tested one-tailed. The order of the videos within the knowledge test has also an effect which can be categorised as small that is only significant if tested one-tailed as well (F (1, 77) = 3.656, p = .060; *partial* η^2 = .045), see Figure 17.
The interaction of the two factors has no significant effect on knowledge acquisition (F (2, 77) = .564, p = .571; *partial* η^2 = .014) as well as the covariate prior declarative knowledge has no effect (F (1, 77) = .000, p = .998; *partial* η^2 = .000).

To find out if the difference in the role model (student vs. professor) or the difference in errors (coping student model vs. mastery student model) can explain the effect on knowledge acquisition, two more ANOVAs are calculated only including one of the factors, respectively and excluding declarative prior knowledge as it has shown to have no effect. The effect on procedural knowledge acquisition can be explained by the video factor errors in example (coping model vs. mastery model) that has a moderate effect (F (1, 47) = 4.112, p = .048; *partial* η^2 = .080) with higher values in knowledge acquisition in the erroneous condition. In contrast, there is no significant effect of the factor role model (student vs. professor) on knowledge acquisition (F (1, 56) = .017, p = .896; *partial* η^2 = .000). Neither the order of videos within the knowledge test, nor the interactions of the factors with the factor order of the videos within the knowledge test are significant.

[26] In a regression the post-values of the knowledge test were predicted by the pre-tests. The residuums are saved and used for further analyses.

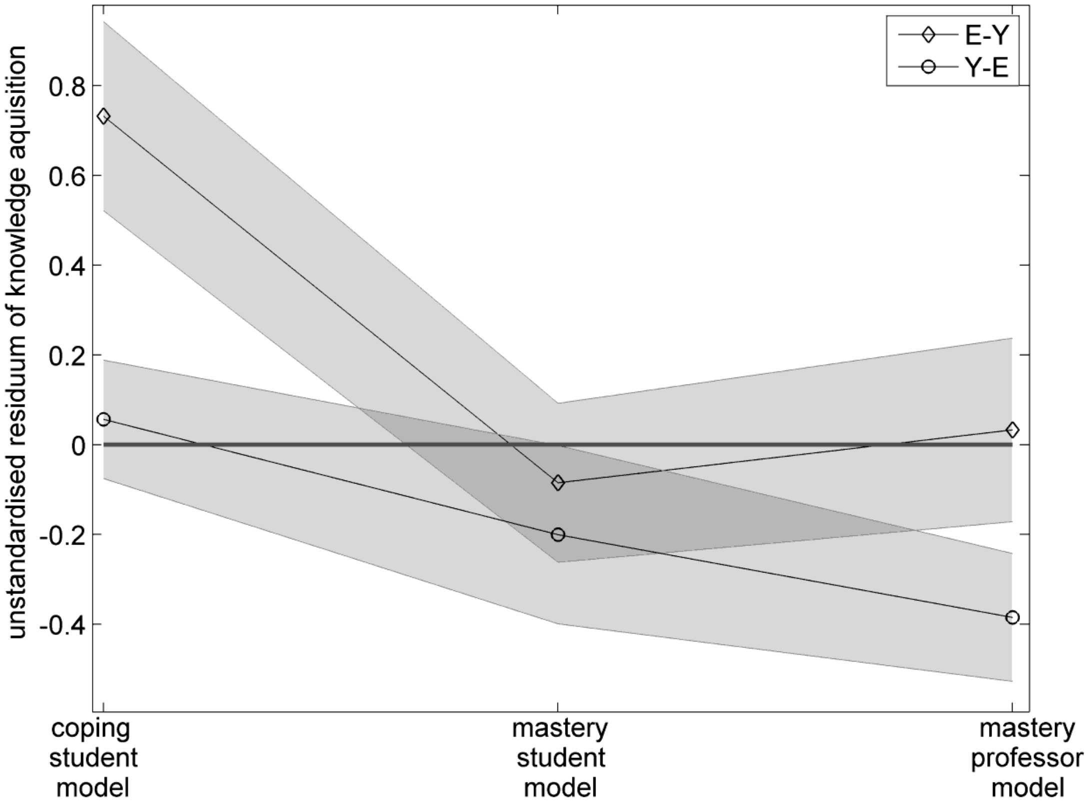

Figure 17: Knowledge acquisition and order of the knowledge-test cases between the video-based modeling worked examples (with standard error).

8.2.3 Relation of Process and Outcome Variables of Role-Model and Errors in Video-Based Modeling Worked Examples

RQ III: To what extent do affect, PSI, and cognitive load mediate the effects of the different video-based worked-examples on performance?

In order to develop a model of learning with video-based modeling worked examples, potentially mediating variables are investigated. Change in affect is examined in RQ III.I and the relation of affect (anxiety, inhibition, and empathy), PSI, cognitive load and learning (self-explanations, performance and knowledge acquisition) is explored by separate bivariate correlations for the factors errors in example and role model (RQ III.II). The information from those analyses are taken into account when one model is tested for each of the factors separately (RQ III.III).

RQ III.I: To what extent do time (pre vs. post) and the factors errors in example and role model have an effect on affect (inhibition and anxiety) when learning with video-based modeling worked examples? Is this effect influenced by the level of Affective PSI with the model?

To analyse the change in anxiety and inhibition and effects of the video-based modeling worked example, two one factorial ANOVAs with repeated measurement with all factor steps of the video-based modeling worked example (coping student model vs. mastery student model vs. mastery professor model) and the repeated measurement of inhibition and anxiety, respectively are made (for the values of affect, see Table 19).

Table 19: Affect before and after the modeling worked example between the factor groups.

		coping student model (n = 21)	mastery student model (n = 26)	mastery professor model (n = 35)
anxiety pre	Mean	30.43	31.15	30.86
	SD	(6.04)	(6.76)	(6.93)
anxiety post	Mean	29.57	27.38	27.23
	SD	(8.37)	(7.76)	(7.55)
		(n = 7)	(n = 18)	(n = 26)
inhibition pre	Mean	22.86	23.83	24.96
	SD	(3.44)	(5.62)	(4.57)
inhibition post	Mean	23.86	22.67	24.77
	SD	(3.93)	(5.27)	(4.90)

There is a significant large effect of repeated measurement / time on anxiety (F (1, 79) = 38.847, p = .000; *Wilk's* Λ = 0.670, *partial* η^2 = .330) but no effect on inhibition (F (1, 48) = 0.072, p = .790; *Wilk's* Λ = 0.999, *partial* η^2 = .001). For anxiety, factor time (pre vs. post) and the factor video-based modeling worked example interact significantly with a medium effect (F (2, 79) = 4.118, p = .020; *Wilk's* Λ = 0.906, *partial* η^2 = .094). Moderate but not significant interaction of the factors for the dependent variable inhibition (F (2, 48) = 1.680, p = .197; *Wilk's* Λ = 0.935, *partial* η^2 = .065). To include Affective PSI in the calculations, two Covariance analyses with factor video-based modeling worked example and Affective PSI as covariate are made for the dependent variable change in anxiety and change in inhibition respectively (for the values of Affective PSI between the factor groups, see Table 14). There is a significant moderate effect of Affective PSI on the change in anxiety (F (1, 77) = 5.664, p = .020, *partial* η^2 = .069) while the effect of the video-based modeling worked example is slightly larger and also significant (F (2, 77) = 3.508, p = .035, *partial* η^2 = .084). Despite slightly more reduction of anxiety for those who worked with the mastery student model, pair-wise comparisons show no significant effects of the factors errors in example (p = .052) and role model (p = .642). For the dependent variable change in inhibition neither the factor video-based modeling worked example (F (2, 46) = 0.484, p = .619, *partial* η^2 = .021), nor the covariate Affective PSI (F (1, 46) = 1.800, p = .186, *partial* η^2 = .038) are significant.

RQ III.II: To what extent are PSI, sympathy, affect, cognitive load, and learning outcomes related?

To analyse relations of mediating variables and learning outcomes, two correlation matrices are compiled separately for the factors errors in example and role model. As previous analyses have shown different effects of the factors errors in example and role-model on the process variables, the correlations will be run separately for both factors. Admittedly, the reduced samples lead to less power of the tests.

First, the relations of the two learning outcomes with the personal factors (affect and empathy) as well as assumed process variables (PSI, cognitive load, and self-explanations) are analysed.

While the learning outcome performance is connected with sympathy for the model (r = .428**, p = .002, n = 49) and extraneous load (r = -.303*, p = .021, n = 58) for those who worked with the student models (coping model vs. mastery model), it is negatively correlated with affect prior to the worked example and after it (inhibition pre and post (r = -.433**, p = .002, n = 47; r = -.432**, p = .002, n = 49) and anxiety pre and post (r = -.249*, p = .044, N = 66; r = -.245*, p = .048, N = 66)), respectively, for those working with the mastery models (student model vs. professor model). In both subgroups the second learning outcome category knowledge acquisition is uncorrelated with performance. For those working with the mastery models (factor role model) cognitive PSI is positively correlated with knowledge acquisition (r = .279*, p = .031, n = 60).

As a next step, correlations between the mediating variables and personal factors are examined.

The supposed mediating variable number of self-explanations is not correlated with any of the variables for the whole sample while the number of correct self-explanations is correlated with the empathy scale Social Concern (r = .302*, p = .033, n = 50) but only for those working with the expert models.

In contrast, there are significant correlations of cognitive load. Intrinsic load ratings show nearly the same patterns for the whole sample with positive correlations of intrinsic load while working with the worked example with post scales of affect (anxiety_errors in example (r = .330*, p = .016, N = 53), inhibition_errors in example (r = .520**, p = .006, n = 26); anxiety_role model (r = .341**, p = .006, n = 64), inhibition_role model (r = .411**, p = .003, n = 49)) and pre scale of anxiety (errors in example (r = .315*, p = .026, n = 50); role model (r = .330**, p = .008, n = 64)). Only the correlation with inhibition prior to the example is limited to those who worked with the mastery models (r = .334*, p = .022, n = 47). Extraneous load is correlated in a same way for the whole sample with negative relations to affective PSI (errors in example (r = -.327*, p = .013, n = 57); role model (r = -.266*, p = .027, n = 69)) while the negative correlation with change in anxiety (r = .321*, p = .032, n = 45) and positive correlation with performance (see the values above) is only reported within the group of learners working with the student models.

Self-assessed empathy only shows correlations based on the subscale *Social Concern*. The ratings are positively correlated for the whole sample with affective PSI (errors in example (r = .432**, p = .002, n = 48); role model (r = .416**, p = .003, n = 49)) and the pre value of inhibition (errors in example (r = .451*, p = .040, n = 21); role model (r = .351*, p = .049, n = 32)). Within the sample of those working with the student models (factor errors in example), *Social Concern* is also positively related to anxiety prior to the worked example (r = .401**, p = .008, n = 43).

For cognitive PSI the correlation pattern is similar for both subgroups but with the factor errors in example, less correlations are significant: It is negatively related to inhibition prior to the worked example (errors in example (r = -.378*, p = .048, n = 28) and role model (r = -

.363*, *p* = .012, *n* = 47)). Inhibition and anxiety after the example are only correlated with cognitive PSI for those working with the mastery models (*r* = -.358*, *p* = .012, *n* = 49; *r* = -.298*, *p* = .015, *n* = 66). Besides correlations with extraneous load and empathy (see above), affective PSI is also correlated with change in anxiety (*r* = -.288*, *p* = .024, *n* = 61) and anxiety prior to the worked example (*r* = .244*, *p* = .048, *n* = 66) but only within the subgroup working with the mastery models.

The rating of the likability of the model is correlated to affective PSI for those working with the student models (*r* = .347*, *p* = .015, *n* = 49) and negatively correlated to the change in inhibition for those working with the expert models (*r* = -.371*, *p* = .022, *n* = 38)

RQ III.III: Is there a model to sufficiently explain the effect of errors and role-model in the video-based modeling worked example on performance?

Two models will be tested separately for the factors errors and role model in the video-based modeling worked example. While a structural equation model would be the first choice for such kind of analysis, the relatively small number of data sets does not allow for this kind of method. Therefore, selective regressions with the variables that proofed to correlate significantly are performed to analyse how the model of factor coping model vs. mastery model works. Performance of the complex skill of BBN is chosen as the targeted outcome variable. The correlation matrices have shown that knowledge acquisition as we measured it is uncorrelated to the performance measurement. Therefore, the aspect of knowledge of the spikes steps is left out in favour of the performance measurement that is closer related to the operationalization of complex skill acquisition.

As a first step, a stepwise regression on the target variable performance with the explanatory variables that have shown to have a direct correlation with the complex skill (likability of the model and extraneous load) is performed with all students included that worked with the video-based worked examples coping model vs. mastery model (*n* = 59).

Both variables are selected for the model which explains 11.9% (corrected R^2, R^2 = .149)[27] of variance of performance ($F(2, 57) = 5.003$, $p = .010$) which is a small to moderate effect. While both explanatory variables account for relevant parts of variance of the performance of complex skill, the *Beta* of likability is larger (*Beta* = .325, *p* = .011) than this of extraneous load (*Beta* = -.259, *p* = .040).

As a second step, a regression on the (unstandardised) residuum of performance that is not explained by extraneous load and likability is performed with four more explanatory variables affective PSI, anxiety and inhibition prior to the worked example, and empathy (social concern). Of all variables only inhibition of the learners prior to the modeling worked example is incorporated in the model as a significant explanatory factor of the residuum of performance ($F(1, 24) = 10.853$, $p = .003$, *Beta* = -.566) and can explain 29.1% (corrected R^2, R^2 = .321)[28] of variance, which is a large effect.

[27] Durbin-Watson-Statistic with a value of 1.761 shows that there should be no auto correlation.
[28] Durbin-Watson-Statistic with a value of 1.999 shows that there should be no auto correlation.

Together, the three predictors can explain 41% of variance of performance within the group of students that worked with the coping model vs. mastery model examples.

As a final step, to find out if the factor errors in example has an additional effect on performance a t-test on the residuum of performance is calculated.
The video-case factor errors in example (coping model vs. mastery model) has no statistically significant effect on the residuum of performance ($t(18) = -1.396$, $p = .180$, $d = -.175$). Because of the reduced sample size the calculation is repeated with the nonparametric Mann-Whitney U-test which confirms the missing effect of the factor errors in example on the residuum of performance of the complex skill ($p = .290$) The effect of the factor errors in example on performance thus can be explained by likability of the model, extraneous load, and inhibition prior to the modeling worked example.

To analyse how the explanatory factors are determined, another regression is run.
As likability as well as extraneous load both correlate significantly with affective PSI a regression with affective PSI as explanatory variable and likability and extraneous load respectively as a target variable is calculated with all students included that worked with the video-based worked examples coping model vs. mastery model and have a performance rating.
Affective PSI is a relevant predictor of both likability of the model and extraneous load. Affective PSI can explain 10.2% (corrected R^2, $R^2 = .121$)[29] of variance of likability which is a small effect ($F(1, 48) = 6.443$, $p = .015$, $Beta = .347$) and 9.1% (corrected R^2, $R^2 = .107$)[30] of variance of extraneous load which is a small effect as well ($F(1, 56) = 6.602$, $p = .013$, $Beta = -.327$).

To analyse how the model of factor role model student vs. professor works, selective regressions with the variables that proofed to correlate significantly are performed.

As a first step, a regression on the target variable performance with the four explanatory variables inhibition and anxiety prior to and after the worked example, respectively, is performed with all students included that worked with the video-based worked examples student model vs. professor model ($n = 49$). Those variables are selected because they showed to correlate with the values of complex skill performance.

Inhibition post is the only factor selected and thus the only significant predictor of performance. It can explain 16.9% (corrected R^2, $R^2 = 18.6$)[31] of variance of performance ($F(1, 48) = 10.763$, $p = .002$, $Beta = -.432$), which is a medium effect. A regression on the residual of the regression of inhibition with cognitive PSI and intrinsic load cannot explain any additional variance.

To analyse how the explanatory factor is determined, another regression is run.
As inhibition post correlates significantly with cognitive PSI and intrinsic load a regression on inhibition post as target variable with cognitive PSI and intrinsic load as explanatory variables

[29] Durbin-Watson-Statistic with a value of 2.023 shows that there should be no auto correlation.
[30] Durbin-Watson-Statistic with a value of 1.789 shows that there should be no auto correlation.
[31] Durbin-Watson-Statistic with a value of 1.894 shows that there should be no auto correlation.

is calculated with all students included that worked with the video-based worked examples student model vs. professor model (n = 80).
Cognitive PSI and intrinsic load experienced during the video-based worked example are both significant predictors of inhibition after the example and can explain 25.1% (corrected R^2, R^2 = 27%)[32] of variance (F(2, 77) = 14.261, p = .000), which is a large effect. While both explanatory variables are significant, intrinsic load has a slightly larger *Beta* = .394 (p = .000) than cognitive PSI (*Beta* = -.319, p = .002). Another regression on the residuum of inhibition with affective PSI which is correlated to cognitive PSI is made. This regression is not significant (F(1, 87) = 1.386, p = .243, *Beta* = .132) and can only explain 0.5% (corrected R^2, R^2 = 1.7%) of variance.

As a final step, to find out if the factor role model in example has an additional effect on performance a t-test on the residuum of performance is calculated.
The video-case factor role model (student model vs. professor model) has a statistically significant large effect on the residuum of performance (t(47) = 2.716, p = .009, d = .796). The effect of the factor errors in example on performance thus cannot be completely explained by cognitive PSI, intrinsic load, and inhibition post and is only slightly reduced from a direct effect size of d = .80 (see 6.2 RQ II of the 2nd study).

8.2.4 Effect of Interventions on Complex Skill Acquisition

Research question IV targets possible effects between the 1st and the 2nd study and the overall effect of the experimental intervention: **Did the intervention of the two studies succeed in improving the course?**

To find out if there is an effect of the round of study on learning, a one-way ANOVA with the factor semester (1st study vs. 2nd study) and two dependent variables (performance of the complex skill and knowledge acquisition) is performed including only the cases of the same condition (incomplete video-based worked example of coping student with prompt).

Table 20: Performance ratings of both studies including those who worked with the incomplete video-based worked example of the coping student without self-explanation prompt.

		1st study (n = 15)	2nd study (n = 28)
performance	Mean	18.67	19.39
	SD	(3.177)	(3.304)
Knowledge acquisition	Mean	-.098	.366
	SD	(1.271)	(.886)
correct self-explanations*	Mean	.13	.68
	SD	(.352)	(.772)

* Significant at the 5% level

There is a small effect of the semester on performance (F (1, 41) = .484, p = .490, *partial* η^2 = .012) and knowledge acquisition (F (1, 36) = 1.754, p = .194, *partial* η^2 = .046) with better

[32] Durbin-Watson-Statistic with a value of 1.359 shows that there should be no auto correlation.

results in the 2nd study. Both effects are not significant. The mean number of correct self-explanations produced by the learners in the 2nd study is significantly higher under the same conditions (F (1, 41) = 6.673, p = .013, *partial* η^2 = .140) but the Levene test shows significant differences of variances (p = .000). Therefore, the nonparametric Mann-Whitney-U-Test is calculated comparing the number of correct self-explanations and supports the finding by rejecting the null hypothesis (p = .015). For all values, see Table 20.

To find out if there is a significant difference between the most effective conditions to foster the complex skill of BBN of both studies respectively, a t-test with factor semester (1st study vs. 2nd study) only including the complete condition of the 1st study and the mastery student model from the 2nd study and the dependent variable performance is made. There is a significant medium effect of the different video-based worked examples on performance with better results of the mastery student model in the 2nd study (t(46) = -2.108, p = .040, d = -0.656), see Figure 18.

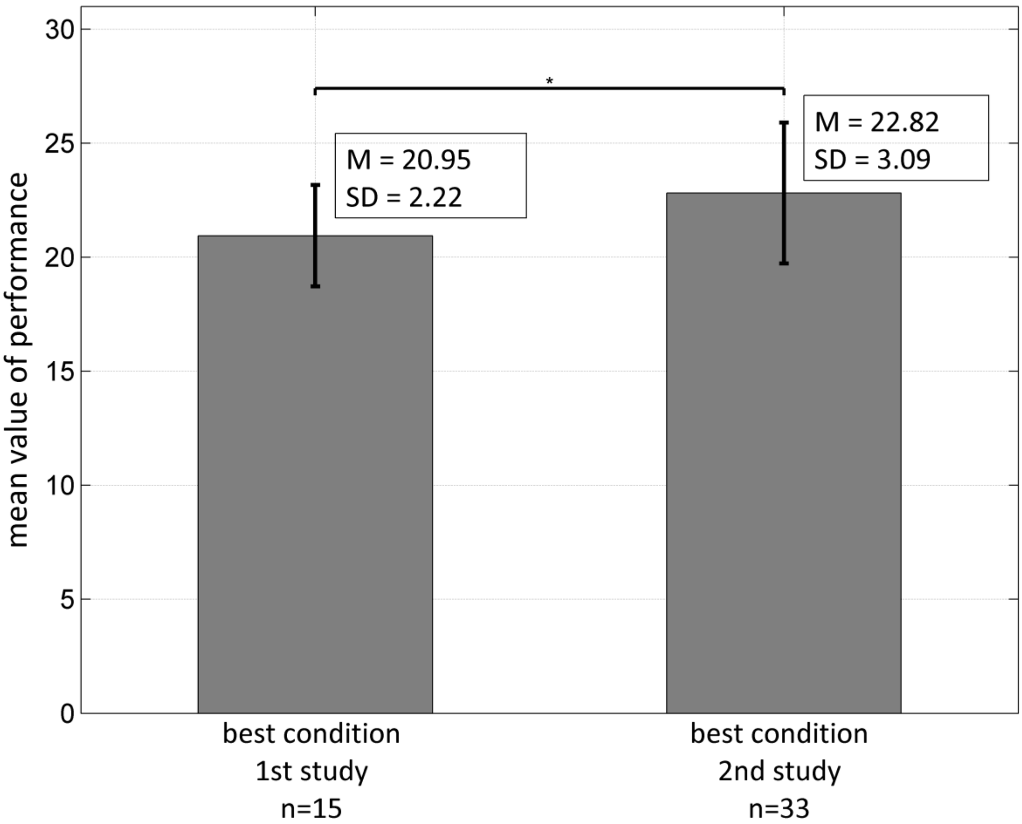

Figure 18: Performance of the students working with the most effective conditions from the 1st study compared to the most effective condition of the 2nd study.

Finally, the control condition of the semester prior to the intervention (M = 17.28, SD = 2.74) will be compared to the most effective condition of the 2nd study (mastery student model, M = 22.82, SD = 3.09) in order to conclude on the total effect size of the introduction of video-based modeling worked examples to the acquisition of the complex skill of BBN. A t-test has been calculated, revealing a significant large effect of the optimized video-based modeling worked example on performance of BBN (t(68) = -7.964, p = .000, d = 1.907).

8.3 Summary and Conclusion on Results from the 2nd Study

The analyses of effects of the factors role model and errors in example are only partly in line with the hypotheses (RQ I). Corresponding to our expectations, the learners experienced more affective PSI with the student model than with the professor model. When both student models were compared, the mastery student was perceived as much more likable compared to the coping student, reaching mean values indicating that he was perceived as ‘likable’ to ‘very likable’ while the latter model was perceived by the observers as relatively neutral in respect to likability. In this case, at least one determinant of sympathy for the respective models seems to be related to their levels of expertise and not - as expected – to resemblance to the learner or authenticity as there are no effects of the role-model on likability.

Also not in line with the hypothesis is the missing effect of the errors in the example on cognitive load.

None of the expected effects on self-explanations were found. There are no effects of the factors role model and errors in example on the self-explanations produced during the video-based worked example. Looking at the descriptives of the number of self-explanations, there are (non-significant) hints that the coping student model condition led to more self-explanations and consequently also more correct ones. This would be in line with recently found positive effects of erroneous examples on self-explanations (Siegler, 2002).

Research question II centered on effects of the experimental variation in the worked examples on learning outcomes. For the first part of this research question the hypothesis can be confirmed: Students learning with the mastery student model example significantly outperformed those who learned from the coping student model. The factor role model also had an effect on performance with better results for the student model. But those effects could only be observed for the dependent variable performance of BBN.

Analyses showed significantly less knowledge acquisition after learning with the mastery model compared to the coping model and no effects of factor role model. Prior declarative knowledge had no effect. This result does not confirm the hypothesis. It was expected that the mastery model would have a positive effect on knowledge acquisition. The missing effect of prior knowledge might be due to the inclusion of the aspect of prior knowledge in the values for knowledge acquisition as those are the residuum of a regression of prior knowledge values on post knowledge values (see 4.4.1).

To adequately answer the question it would be suitable to include prior experience as an additional factor or covariate. Unfortunately, the number of participants per cell was very low and in one case only consisted of one learner if prior experience of BBN[33] was included. This is, besides the three-factorial design, due to the fact that most of the students indicated to be new to the field. Therefore, the factor prior experience was replaced by the covariate prior declarative knowledge. Anyway, future studies with larger samples should include the factor prior experience of the actual performance in the design.

---

[33] *‚Ich habe bereits eine schwierige Nachricht übermittelt.‘* (Yes vs. No)

Anxiety could generally be reduced after the worked example compared to the values prior to it (RQ III.I) which is in line with the hypothesis and a replication of the result in the 1st study. There is also a main effect of the video-based modeling worked example which is remained when affective PSI is added as a covariate that has also a significant effect on the reduction of anxiety with higher values of affective PSI related to more reduction of anxiety, confirming the hypothesis. Comparing the different worked examples, there is a hint towards a positive effect of the mastery student model compared to the coping student model on the reduction of anxiety related to BBN, but no significant effect is measurable. In contrast, inhibition is not generally reduced which contradicts results from the 1st study. On the other hand a direct comparison is not possible as the scales for both aspects of affect were extended.

The second part of this research question (RQ III.II) targets two models that were tested to explain the effects of the factors role model and errors in the video-based modeling example on performance. After a correlation of all relevant outcome variables and process variables separately for both factors that revealed some different patterns, two models were generated for testing.

Only the model for errors in example with the coping student model and the mastery student model could be confirmed with mediation by PSI (likability of the model and affective PSI) and extraneous load influenced by the modeling worked example and additional effects by the learners' degree of inhibition to perform the complex skill before the learning intervention.

It remains open which factors are responsible for the effect of the role model on performance (RQ III.III) as only the post value of inhibition, intrinsic load, and cognitive PSI can explain essential variance of performance. But neither inhibition and cognitive PSI nor intrinsic load explain the large effect of the factor role model on performance.

Generally, a structural equation model including all of the factors that are expected to influence the acquisition of the complex skill of BBN would have been the best way to test the two models, respectively. However, the data of this study do not match the criteria to analyse such a model as it is recommended to include $N \geq 100$ or better $N \geq 200$ or at least five times as many as parameters to be estimated (Backhaus, Erichson, Plinke, & Weiber, 2006). Future studies should thus collect data on a larger scale in order to analyse the expected models without the risk of alpha error inflation.

The final research question IV links both studies and compares effects of the same condition between the rounds of data collection and the most effective conditions, respectively. Contradicting the hypothesis that the same condition in both studies should also have the same effect on learning, small to medium positive effects of the 2nd study were revealed. The only effect that is significant though is the higher number of correct self-explanations in the 2nd study. As the student sample should be comparable, the slightly better results under the same condition could be due to the fact that the text on the spikes the students read before working with the modeling example has been translated to German for the 2nd study. A second analysis comparing both rounds of the study only included those conditions that had shown to have the best effects on learning, revealed a moderate positive effect of the incomplete mastery student model example (best condition of the 2nd study) over the complete example of the coping student model (best condition of the 1st study). The supremacy of the best condition of the 2nd study over this of the 1st study in respect to performance of BBN confirms the hypothesis that a mastery model making no or at least only little mistakes leads to more acquisition of a complex skill than an erroneous example consisting of a coping model.

The final analysis comparing the control condition where students learned without video-based modeling worked examples to the most effective condition of the 2nd study revealed a significant large effect, underlining the success of the intervention.

9 Discussion

In this chapter, the results of both studies are summarised and discussed against the empirical background. Subsequently limitations and future research is pointed out before the final conclusion is drawn.

9.1 Summary of Results

This thesis centres on the mechanisms of the combination of modelling and worked examples and how examples of this instructional approach should be designed to foster complex skill acquisition. In particular, three aspects were investigated: (a) How should video-based modeling worked examples be constructed structure wise to support the acquisition of a complex skill?; (b) What kind of model leads to the best outcome performance of the learners?; and (c) What are the underlying mechanisms involved?
Also, the aspect of affect and learning with video-based modeling worked examples has been investigated, adding the aspect of (d) What is the role of negative affect and empathy in the learning process with video-based modeling worked examples?

The questions have been targeted with two studies within the same population of medical students and the following conclusions can be drawn: (1) The rules for successful worked examples cannot be completely translated to erroneous video-based worked examples for complex skill acquisition. There was a negative effect of incompleteness of the video-based worked example on knowledge acquisition and performance of BBN during the 1st study. However, the number of correct self-explanations was higher for those learners who were supported with explanation prompts within the erroneous worked example, adding knowledge to the question: 'Which kind of self-explanation prompt supports the elicitation of learning?' Admittedly, other factors are relevant for the production of correct self-explanations as well as the amount thereof being higher during the 2nd data collection when the students received a German version of the knowledge input (spikes text). Furthermore, the number of correct self-explanations was not related to learning outcomes which contradicts research on worked examples. (2) When different models embedded in video-based worked examples were compared, the student model proved to be more effective than a professor model, supporting findings from research on observational learning that recommends models that are equal to the learners in respect to appearance. However, this is only true if the student is an expert in the displayed competence in contrast to a coping model displaying erroneous behaviour, shedding light on an important differentiation. So far, expert models were often put on the same level as experienced adult persons (Baum & Gray, 1992) while in this study two experts of different status were compared showing positive effects of student experts. In this scope it has further been confirmed that the likability of a model is connected to better learning results (Bandura, 1986) but it has also been shown that this likability is strongly affected by the level of competence displayed, which is a novel finding. Those results cannot confirm positive effects of erroneous examples as reported by research from worked examples (Stark et al., 2011). (3) Part of the variance of the positive effect of the mastery student model compared to the coping student model (erroneous example) on performance can be explained by the likability of the model and affective PSI with the model. Those results are in line with findings from research on effective models for observational learning (van Ophuysen & Hannover, 2005). In addition, some light is shed on the factors that lead to the categorisation of an observed model as likable. Furthermore, the theory of PSI has shown to be a relevant concept in the scope of example-based learning. The sensation of extraneous load and initial inhibition to perform the complex skill is hindering, confirming previous empirical results.

The mechanism behind the positive effect of the student model over the professor model could not be clarified by the study at hand. While part of the correlation pattern of mediating variables is the same as for the comparison of the student models, the only factor that explains a significant amount of variance of the performance of BBN besides the factor role model (student model vs. professor model) is a negative effect of inhibition experienced by the learners after the worked example. While intrinsic load and the cognitive scale of PSI can explain part of the variance of inhibition, it remains unclear how exactly this affective state is caused and how it mediates the positive effect of the student role model. Generally, those students who took their time to deliver the bad news to the patients performed better. (4) Anxiety of BBN was reduced across all instructional conditions independently from structure, errors included and role model displayed, even though the learners indicated that their levels of anxiety prior to the worked example were rather low. This finding argues against findings from other simulations in medical education that were shown to enhance anxiety levels in the learners (Paskins & Peile, 2010). Furthermore, it has been confirmed that anxiety as well as the empathy-subscale describing the trait of *Social Concern* play a role in the learning process (see: model for learning with vs. without errors in example), by accounting for part of the variance of affective PSI which in turn explains variance of performance. Surprisingly, the second empathy-scale *Readiness for Empathy* that measures the 'ability and willingness to put oneself in the position of a fictional other in respect to experience and behaviour' is not related to any mediating or outcome variable. In contrast to the latter scale, *Social Concern* measures empathic behaviour in real situations. Even though (self-assessed) empathy is positively related to the quality of BBN, the effect is rather small, indicating that being empathic is not enough and even those people need training. A reduction of inhibition to break bad news could only be found during the 1st study, preventing the author's ability to make a clear statement on this matter.

9.2 Limitations and Future Research

Both studies have been carried out in a real learning context with a limited number of participants. This restriction was on the one hand a limiting factor for the number of factor groups to be compared resulting in two rounds of data collection, and on the other hand the controlled field-study character of the whole study ensures results that are externally valid and provide grounds for optimism in respect to successful application in similar academic contexts.

There are some downsides, however, of the very applied and curriculum-embedded data collection: Both studies were carried out in the context of a mandatory course embedded in the curriculum of medical education. The restrictions on time for the learning unit did not allow for more than one complete case to be observed, albeit modelling has been shown to yield better results in respect to accuracy of reproduction and identification with the model of repeated exposition to it (Bandura, 1986; Carroll & Bandura, 1990). I tried to compensate for this flaw with the possibility to repeatedly watch single episodes of the modelling example and measure PSI to include the relation to the model in the evaluation. Future studies should adhere to the measurement of PSI or similar in order to control the interpersonal aspects of model and observer, to add more examples and a repeated training session after 30-34 months to counteract initial skills decline (Wood, 2010). Furthermore, the desire for collaborative work expressed by the students should be implemented, also to investigate effects on the perception of PSI. The change in the external form of the theoretical input on the spikes (paper vs. condensed English version vs. condensed German version) might account for some variance of learning between the three data sets and cannot be controlled in the study at hand. This variance is an effect of the field-study character of the studies: as the goal of the cooperation partners (surgical clinics) was to enhance the quality of the teaching, the students' wish (*here*: translate text to German) had to be implemented immediately.

To better compare both studies, a complete accordance of all scales that were included, respectively, would have allowed for more comparisons. However, there were reasons for changing the scales for cognitive load and affect (anxiety and inhibition). In contrast to the 1st round of data collection, cognitive load has been shown to play a role in learning with video-based modeling worked examples and explain variance in its effect on performance of a complex skill. The revelation of effects could potentially be due to the changes in measurement. The items applied in the 2nd study are in accordance with findings in cognitive load research that recommend measuring extraneous load and intrinsic load and refraining from scales for germane load as this cannot be clearly distinguished from the operationalisation of intrinsic load (Kalyuga, 2011; Leppink et al., 2013; Leppink et al., 2014).

Another comparison which is not feasible with the data available is the kind of self-explanations of the learners from complete and incomplete conditions. Unfortunately, there was no possibility for those working with complete examples to express how they explained the solution steps of the worked example. This should be included in future studies and might help to explain the better learning results in the complete condition compared to the incomplete condition in the 1st study.

The measurement of declarative knowledge has been the same for the whole study and in both rounds of data collection, the post-test values were reduced. Bearing in mind that some learners reached close to the maximum value in the pre-test, this test might have been too easy to solve. Consequences are, on the one hand, to enhance difficulty of knowledge acquisition tests in future studies. On the other hand, this phenomenon sheds light on the important differentiation of declarative knowledge acquired and the transfer to performance, once again supporting the relevance of performance-based authentic assessments to gain valid insight into the acquisition of complex skills instead of inert declarative knowledge.

The model for learning with vs. without errors in a video-based modeling worked example has been successfully explained by the variables affective PSI, likability of the model, and extraneous load, generally exhibiting positive effects of the mastery model (without errors). This main effect contradicts findings from empirical worked example research indicating positive effects of erroneous examples. However, there are studies suggesting that there are more crucial personal characteristics that have an effect on the management of errors. Positive effects were found for openness to experience and agreeableness as well as ability in the field (Loh, Andrews, Hesketh, & Griffin, 2013). Therefore, future studies have to investigate the effects of erroneous examples in respect to personality traits and states that may moderate the beneficial vs. detrimental effects. Furthermore, the approach to the errors displayed may be critical. A study in the field of error management training has proven positive effects of a foregone positive error framing on metacognition and self-efficacy (Steele-Johnson & Kalinoski, 2014). Explicitly telling the learners that errors are useful may result in positive effects of erroneous worked examples.

The effect of errors in a modeling worked example on the perceived likability of the model should be investigated further. In this study the mastery student model was perceived as much more likable than the coping student model. As the conditions did not include the same person as model, in future studies the same model should be displayed in both conditions (with vs. without errors) to rule out effects of other model related characteristics. Generally, the models should be tested with structural equation modelling based on larger samples.

Albeit the positive effect of all video-based modeling worked examples to reduce anxiety of BBN and the inhibition to deliver them, standard deviations are rather large. This is an indicator for a diverse picture of the affective state between the learners which should be investigated further in more depth. Ratings of the participants' empathic behaviour during BBN were not included in the analyses because the assessment by the SPs did not appear to be objective and reliable. The limitation of data sources to only the self-assessment ratings by the learners themselves is of limited quality. In a follow-up study the SPs should be trained beforehand to be able to make ratings that match quality criterions.

Another aspect of learning with SPs is the empirical result that performance is better with SPs than with real patients (Pieters, Touw-Otten, & De Melker, 1994). This finding highlights the difference between competence as assessed in a testing situation and true performance in real clinical settings. On the other hand it has been shown that when an assessment is perceived as more authentic the studying becomes deeper and more professional skills are being developed (Gulikers et al., 2008). So even if there is a difference between simulated BBN and BBN in real life situations the assessment with SPs is still superior to less authentic assessments.
Results from the 1st study indicated a positive effect of meeting and performing with the same person previously seen in a video-based modeling worked example. This effect was not replicated with the data from the 2nd study. Still, those findings might point towards limitations in respect to transfer of the results. Meeting and performing with the same SP as seen in the video can be categorized as less transfer than applying the complex skill of BBN to the conversation with another SP. Future studies should therefore avoid the recurrence of the SP from an example in a follow-up face-to-face situation if the latter is the data collection of complex skill acquisition.
In terms of external validity the numerous SPs included in the study can be seen as an enhancement of transferability of the results to real life doctor patient settings. At the same time the diversity of SPs in respect to their pleasantness can be categorised as a confounding aspect. Also, the increased difficulty for the students in performing with those patients, we did not find any effects of the different SPs and also no interaction effect of the SP shown in the learning material and the SP the students met afterwards and delivered the bad news to. It also did not have an effect on the performance if the students met a standardised patient of the same gender or not.
To further investigate how the SP needs to be designed to be most helpful, further studies focusing on systematically varying the SPs characteristics will have to be performed.

The limited effects of the learners' empathy might be partly explained by the division of empathy made by Leibetseder and colleagues (2007) into a spontaneous emotional reaction and an 'active volitional sensitivity'. While many students advanced the opinion that the skill of BBN was a stable given competence that *one possesses or not* our results show that we should rather support people in finding constructive ways to benefit from empathy as requested by Leibetseder and colleagues (2001). To add more insights into the role of affective PSI, affect, and empathy caused by abstract modelling a variation of how many emotions the models show may be implemented.

For the study we did not develop an own rating scheme for good communication but based the training content on a pre-existing and well established six-step scheme (Baile et al., 2000). The results are therefore limited to the competence of applying this communication scheme and not communication skills in general. In future studies, it would be interesting to have experts from different fields (like doctors who have to break bad news, patients and relatives of affected patients) rate the videos of the BBN conversations to analyse the correlation of their evaluation and the one we applied on the basis of the spikes. Also the behaviour of the different SPs (from the learning material as well as in the simulated performance test) should be analysed further as communication always depends on all participating parties (Realdon et al., 2006). The spikes were used as a structuring element for the video-based worked example. This is not the best possible solution in respect to optimal worked example design as Shen and Tsai (2009) state in their review that sub goals are superior to steps (process principle).

9.3 Final Conclusion

Taken together, those conclusions underline the assumption that abstract modelling with video-based worked examples support the acquisition of the complex skill of BBN. The two consecutive applied field-studies have thus led to an improvement of medical education by the introduction of video-based modeling worked examples. The process of design-based research led to advanced learning material based on theoretical considerations taken from research of example-based learning in the domain of worked examples as well as observational learning.

The SPs support this empirical finding by giving very positive feedback to the instructor, claiming much better performances by the students of both studies compared to the semester before the intervention was initiated. This adds to the validity of the coding scheme applied for performance of the complex skill of BBN. Those effects are unrelated to knowledge acquisition which has not been fostered as well. This picture has already been described by van Gog and Rummel (2010) and has even been reported by earlier studies that observed missing effects on recall of the spikes steps while at the same time the learners reported amelioration of confidence in their own performance (Bonnaud-Antignac et al., 2010). As the learning goal goes beyond good scores in multiple choice tests, knowledge tests are less relevant than more practical tests. At the LMU, OSCE stations have been established and depict the development towards more and more practical oriented tests/tasks. Therefore, the analyses targeting the learners' actual behaviour in the BBN conversations can be categorised as more relevant and significant.

When comparing the mean performance of the control group from the semester before the intervention began with the most effective condition of the 2nd study (mastery student model), the effect on performance is compellingly strong. Between the two rounds of data collection, the final comparison of the best conditions from both studies reveals the supremacy of the incomplete video-based examples with the mastery student model from the 2nd study over the complete worked examples showing the coping student model.

We have also been successful in gaining some insights into the moderation of affect and cognitive load. Anxiety of BBN could effectively be reduced by working with the video-based modeling worked example, contradicting findings from Paskins and Peile (2010) from clinical settings. The largest reduction was experienced after working with a worked example showing a mastery model which might hint at the save and supporting character of the learning environment as a reason behind reducing anxiety in the students. Anxiety is closely interlinked with affective PSI which supports the inclusion of the latter concept in research on video-based modeling worked examples. Opposing former empirical research (Bauer et al., 1983; Sarason et al., 1968; Turner et al., 1974; Zeidner, 1998; Zeidner, 2007) no correlation of affect and the learning outcome was found. Affective PSI, on the other hand, can explain variance in skill acquisition for the model of errors in video-based examples and is positively related to the reduction of anxiety. Results on this relation, which has previously not been analysed, provide a clearer picture of the mechanisms of anxiety during learning with modeled examples and must be investigated further. Inhibition as the second affective aspect that has been studied remains less informative. Different results between the two studies cannot be compared due to a change in items on the scale and the expected negative effect of inhibition on the complex skill as reported by Parle et al. (1997) has not been replicated. Before future studies investigate this variable the scale for inhibition must be validated.

In respect to cognitive load, a direct effect of errors in example or role model on the perception of extraneous load was missing. Still, regressions revealed that extraneous load and likability of the model presented in the worked example account for a relevant amount of variance in the outcome performance of the complex skill of BBN when learning with the worked examples presenting a student model. Further relevant factors for the positive effect of the expert student model are again affective PSI as well as inhibition that the learners experienced towards BBN before the learning intervention. This adds the perceived likability of the model as another central variable. However, it must be said that it is an open question what other factors influence the perception of a model as likable besides affective PSI and its level of competence.

The positive effect of the student role model compared to the professor model cannot be explained satisfactorily. However, intrinsic load and cognitive PSI in the learning process as well as the inhibition to BBN are relevant factors in explaining the positive effect of the mastery student role model on the acquisition of the complex skill of BBN.

While students favour the embedment of videos in example-based learning it has to be concluded from the findings of my studies that it is not sufficient to exclusively base the conception of the learning material on findings from research on worked examples. Structure-wise incomplete video-based modeling examples did have a negative effect on the learning outcome which is contradictory to empirical findings from worked examples research (Renkl et al., 2002; Shen & Tsai, 2009; Stark, 1999) but might be confounded with an unbalanced distribution of the SPs to the conditions. While there was a positive effect of the explanation prompts on the number of correct self-explanations during the video-based modeling example, no correlation of the latter on the learning outcome was found. This finding is not opposed to findings for self-explaining during worked-examples as the empirical results are inconclusive (Berthold & Renkl, 2009; Chi et al., 1989; Renkl, 2005; Schworm & Renkl, 2006). Looking at effects of the content of video-based examples the positive effect of a correct modeling example on complex skill acquisition is in accordance with empirical findings from worked examples research (Große & Renkl, 2007) and observational learning (Bandura, 1986; Baum & Gray, 1992; Boekhout et al., 2010). The positive effect of the erroneous example on knowledge acquisition is in line with findings from Stark et al. (2011) on learning with erroneous worked examples. Those results show that the effects of video-based modeling examples and classic worked examples are comparable in respect to their content while recommendations in respect to structure cannot be transferred directly. As videos appear in principle quite different from written text, which is the format of most worked examples included in empirical research, this result does not come by complete surprise. Aspects related to the model embedded in the example have shown to be more relevant when learning with video-based modeling worked examples. Still, the mechanisms behind learning from video-based modeling examples in respect to self-explanations must be investigated in future studies especially by using think aloud protocols and comparisons also during complete examples.

For the design of video-based modeling examples, some recommendations can be drawn from the studies presented. The selection of the model as well as the quality of the displayed skill must be chosen by considering the likability of the model, its authenticity to the learner, and the prestige of the role-model. For the training of complex skills related to negative emotions it should thus be aimed at modeling examples that induce high affective PSI. Concluding from our results those are models showing correct behaviour while being no expert model in the traditional sense of the term[34] but rather resemble the learner. Empathy of the learner seems to be less relevant when learning with modeling worked examples than PSI experienced with the models. Still, effects of empathy based on data assessed by others than the learners themselves should be investigated in future research.

The positive effects of video-based modeling worked examples on the acquisition of complex skills have been found in the domain of medical education but must not be limited to it. The application of this approach to other fields of learning is promising and should be tested in future studies.

[34] Traditionally, expert models are often represented by experienced representatives of a certain profession and of an average age.

References

Aderhold. (2008). *Wie Kinder mit Lösungsbeispielen lernen*. Saarbrücken: VDM Verlag Dr. Müller.

Anderson, T., & Shattuck, J. (2012). Design-based research: A decade of progress in education research? *Educational Researcher, 41*(1), 16-25.

Anolli, L., Vescovo, A., Agliati, A., Mantovani, F., & Zurloni, V. (2006). Simulation-based training of communication and emotional competence for the improvement of physician-patient relationship. *Annual Review of CyberTherapy and Telemedicine, 4*, 79-86.

Antonenko, P., Paas, F., Grabner, R., & van Gog, T. (2010). Using Electroencephalography to Measure Cognitive Load. *Educational Psychology Review, 22*(4), 425-438. doi: 10.1007/s10648-010-9130-y

Atkinson, R. K., Derry, S., J. , Renkl, A., & Wortham, D. (2000). Learning from Examples: Instructional Principles from the Worked Examples Research. *Review of Educational Research, 70*(2), 181-214.

Backhaus, K., Erichson, B., Plinke, W., & Weiber, R. (2006). *Multivariate Analysemethoden* (11th Edition ed.). Berlin: Springer.

Baile, W. F., Buckman, R., Lenzi, R., Glober, G., Beale, E. A., & Kudelka, A. P. (2000). SPIKES-A six-step protocol for delivering bad news: application to the patient with cancer. *The Oncologist, 5*(4), 302-311.

Baile, W. F., Lenzi, R., Kudelka, A. P., Maguire, P., Novack, D., Goldstein, M., . . . Bast, R. C. (1997). Improving physician—patient communication in cancer care: Outcome of a workshop for oncologists. *Journal of Cancer Education, 12*(3), 166-173. doi: 10.1080/08858199709528481

Bandura, A. (1965). Behavioral modification through modeling procedures. In L. Krasner & L. Ullmann, P. (Eds.), *Research in Behavioral Modification* (pp. 310-393). New York: Holt, Rinehart & Winston, Inc.

Bandura, A. (1969). *Principles of behavior modification*. New York: Holt, Rhinehart, & Winston.

Bandura, A. (1977). *Social learning theory*. Stanford University: Prentice-hall, Inc, Englewood Cliffs.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ US: Prentice-Hall, Inc.

Bauer, G. P., Schlottmann, R. S., Bates, J. V., & Masters, M. A. (1983). Effect of state and trait anxiety and prestige of model on imitation. *Psychological Reports, 52*(2), 375-382.

Baum, B. E., & Gray, J. J. (1992). Expert modeling, self-observation using videotape, and acquisition of basic therapy skills. *Professional Psychology: Research and Practice, 23*(3), 220-225.

Berthold, K., & Renkl, A. (2009). Instructional Aids to Support a Conceptual Understanding of Multiple Representations. *Journal of Educational Psychology, 101*(1), 70-87. doi: 10.1037/a0013247

Berthold, K., Röder, H., Knörzer, D., Kessler, W., & Renkl, A. (2011). The double-edged effects of explanation prompts. [Die zweischneidigen Effekte von Erklärungsaufforderungen]. *Computers in Human Behavior, 27*(1), 69-75. doi: 10.1016/j.chb.2010.05.025

Bielaczyc, K., Pirolli, P. L., & Brown, A. L. (1995). Training in Self-Explanation and Self-Regulation Strategies: Investigating the Effects of Knowledge Acquisition Activities on Problem Solving. *Cognition and Instruction, 13*(2), 221-252. doi: 10.2307/3233714

Blandin, Y., Lhuisset, L., & Proteau, L. (1999). Cognitive processes underlying observational learning of motor skills. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 52A*(4), 957-979.

Blandin, Y., & Proteau, L. (2000). On the cognitive basis of observational learning: Development of mechanisms for the detection and correction of errors. *Quarterly Journal of Experimental Psychology: Section A, 53*(3), 846-867. doi: 10.1080/027249800410571

Boekhout, P., Van Gog, T., Van De Wiel, M. W. J., Gerards-Last, D., & Geraets, J. (2010). Example-based learning: Effects of model expertise in relation to student expertise. *British journal of educational psychology, 80*(4), 557-566.

Bonnaud-Antignac, A., Campion, L., Pottier, P., & Supiot, S. (2010). Videotaped simulated interviews to improve medical students' skills in disclosing a diagnosis of cancer. *Psycho-Oncology, 19*(9), 975-981. doi: 10.1002/pon.1649

Booth, J. L., Lange, K. E., Koedinger, K. R., & Newton, K. J. (2013). Using example problems to improve student learning in algebra: Differentiating between correct and incorrect examples. *Learning and Instruction, 25*(0), 24-34. doi: http://dx.doi.org/10.1016/j.learninstruc.2012.11.002

Carroll, W. R., & Bandura, A. (1990). Representational guidance of action production in observational learning: A causal analysis. *Journal of Motor Behavior, 22*(1), 85-97.

Catrambone, R. (1994). Improving examples to improve transfer to novel problems. *Memory & Cognition, 22*(5), 606-615. doi: 10.3758/BF03198399

Chi, M. T. H. (2000). Self-explaining expository texts: The dual process of generating inferences and repairing mental models. In R. Glaser (Ed.), *Advances in instructional psychology: Educational design and cognitive science* (pp. 161-238). Mahwah, NJ: Erlbaum.

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science: A Multidisciplinary Journal, 13*(2), 145-182.

Chi, M. T. H., de Leeuw, N., Chiu, M.-H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*, 439-477.

Choi, H.-H., van Merriënboer, J. J. G., & Paas, F. (2014). Effects of the physical environment on cognitive load and learning: Towards a new model of cognitive load. *Educational Psychology Review*. doi: 10.1007/s10648-014-9262-6

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale: Lawrence Erlbaum Associates.

Cook, D. A., Erwin, P. J., & Triola, M. M. (2010). Computerized virtual patients in health professions education: A systematic review and meta-analysis. *Academic Medicine, 85*(10), 1589-1602.

Cooper, G., & Sweller, J. (1987). Effects of schema acquisition and rule automation on mathematical problem-solving transfer. *Journal of Educational Psychology, 79*(4), 347-362. doi: 10.1037/0022-0663.79.4.347

Cox, R., McKendree, J., Tobin, R., Lee, J., & Mayes, T. (1999). Vicarious learning from dialogue and discourse. *Instructional Science, 27*(6), 431-458. doi: 10.1023/A:1003489631631

Crippen, K. J., & Earl, B. L. (2007). The impact of web-based worked examples and self-explanation on performance, problem solving, and self-efficacy. *Computers & Education, 49*(3), 809-821.

Curry, L. A. (2004). *The effects of self-explanations of correct and incorrect solutions on algebra problem-solving performance.* Paper presented at the Annual Conference of the Cognitive Science Society, Chicago.

Daniels, L. M., Stupnisky, R. H., Pekrun, R., Haynes, T. L., Perry, R. P., & Newall, N. E. (2009). A longitudinal analysis of achievement goals: From affective antecedents to emotional effects and achievement outcomes. *Journal of Educational Psychology, 101*(4), 948-963.

De Bock, D., Verschaffel, L., Janssens, D., Van Dooren, W., & Claes, K. (2003). Do realistic contexts and graphical representations always have a beneficial impact on students' performance? Negative evidence from a study on modelling non-linear geometry problems. *Learning and Instruction, 13*(4), 441-463.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic Motivation and Self-Determination in Human Behaviour.* New York, NY: Plenum Press.

Decker, P. J. (1980). Effects of Symbolic Coding and Rehearsal in Behavior-Modeling Training. *Journal of Applied Psychology, 65*(6), 627-634.

Decker, P. J. (1984). Effects of different symbolic coding stimuli in behavior modeling training. *Personnel Psychology, 37*(4), 711-720.

DeLeeuw, K. E., & Mayer, R. E. (2008). A comparison of three measures of cognitive load: Evidence for separable measures of intrinsic, extraneous, and germane load. *Journal of Educational Psychology, 100*(1), 223-234.

Dibble, J. L., & Rosaen, S. F. (2011). Parasocial interaction as more than friendship: Evidence for parasocial interactions with disliked media figures. *Journal of Media Psychology: Theories, Methods, and Applications, 23*(3), 122-132.

Dowaliby, F., & Schumer, H. (1973). Teacher-centered versus student-centered mode of college classroom instruction as related to manifest anxiety. *Journal of Educational Psychology, 64*. doi: 10.1037/h0034590

Dunning, D., Johnson, K., Erlinger, J., & Kruger, J. (2003). Why People Fail to Recognize Their Own Incompetence. *Current Directions in Psychological Science, 12*(3), 83-87.

Dymond, R. F. (1949). A scale for the measurement of empathic ability. *Journal of Consulting Psychology, 13*(2), 127-133. doi: 10.1037/h0061728

Ertelt, A. (2007). *On-Screen Videos as an Effective Learning Tool. The Effect of Instructional Design Variants and Practice on Learning Achievements, Retention, Transfer, and Motivation.* Albert-Ludwigs Universität, Freiburg im Breisgau.

Evans, A. W., Leeson, R. M. A., & Petrie, A. (2007). Reliability of peer and self-assessment scores compared with trainers' scores following third molar surgery. *Medical Education, 41*(9), 866-872.

Fallowfield, L., & Jenkins, V. (2004). Communicating sad, bad, and difficult news in medicine. *The Lancet, 363*(9405), 312-319.

Fischbeck, S., Mauch, M., Leschnik, E., Beutel, M. E., & Laubach, W. (2011). Überprüfung ärztlicher kommunikativer kompetenz mittels einer OSCE bei studierenden der medizin im ersten studienjahr. = Assessment of communication skills with an OSCE among first year medical students. *PPmP: Psychotherapie Psychosomatik Medizinische Psychologie, 61*(11), 465-471.

Fischer, K. W. (1980). A theory of cognitive development: The control and construction of hierarchies of skills. *Psychological Review, 87*(6), 477-531. doi: 10.1037/0033-295X.87.6.477

Gentile, D. A., Anderson, C. A., Yukawa, S., Ihori, N., Saleem, M., Ming, L. K., . . . Sakamoto, A. (2009). The effects of prosocial video games on prosocial behaviors: International evidence from correlational, longitudinal, and experimental studies. *Personality and Social Psychology Bulletin, 35*(6), 752-763.

Gerjets, P., Scheiter, K., & Catrambone, R. (2006). Can learning from molar and modular worked examples be enhanced by providing instructional explanations and prompting self-explanations? *Learning and Instruction, 16*(2), 104-121.

Giles, D. C. (2002). Parasocial interaction: A review of the literature and a model for future research. *Media Psychology, 4*(3), 279-305.

Ginns, P., Chandler, P., & Sweller, J. (2003). When imagining information is effective. *Contemporary Educational Psychology, 28*(2), 229-251. doi: 10.1016/S0361-476X(02)00016-4

Greitemeyer, T., & Osswald, S. (2010). Effects of prosocial video games on prosocial behavior. *Journal of Personality and Social Psychology, 98*(2), 211-221.

Greitemeyer, T., Osswald, S., & Brauer, M. (2010). Playing prosocial video games increases empathy and decreases schadenfreude. *Emotion, 10*(6), 796-802.

Große, C. S., & Renkl, A. (2007). Finding and fixing errors in worked examples: Can this foster learning outcomes? *Learning and Instruction, 17*(6), 612-634.

Gulbrandsen, P., Madsen, H. B., Benth, J. S., & Lærum, E. (2010). Health care providers communicate less well with patients with chronic low back pain – A study of encounters at a back pain clinic in Denmark. *PAIN, 150*(3), 458-461.

Gulikers, J. T. M. (2006). *Authenticity is in the Eye of the Beholder. Beliefs and perceptions of authentic assessment and the impact on student learning* (doctoral), Open University of the netherlands, Heerlen.

Gulikers, J. T. M., Bastiaens, T. J., & Kirschner, P. A. (2004). A Five-Dimensional Framework for Authentic Assessment. *Educational Technology Research and Development, 52*(3), 67-86.

Gulikers, J. T. M., Kester, L., Kirschner, P. A., & Bastiaens, T. J. (2008). The effect of practical experience on perceptions of assessment authenticity, study approach, and learning outcomes. *Learning and Instruction, 18*(2), 172-186.

Gunawardena, C. N. (1995). Social presence theory and implications for interaction and collaborative learning in computer conferences. *International journal of educational telecommunications, 2/3*(1), 147-166.

Hanna, M., & Fins, J. J. (2006). Viewpoint: Power and Communication: Why Simulation Training Ought to Be Complemented by Experiential and Humanist Learning. *Academic Medicine, 81*(3), 265-270. doi: 10.1097/00001888-200603000-00016

Harden, R. M., Stevenson, M., Wilson Downie, W., & Wilson, G. M. (1975). Assessment of Clinical Competence using Objective Structured Examination. *British Medical Journal*(1), 447-451.

Hartmann, T. (2008). Parasocial interactions and paracommunication with new media characters. In E. Konijn, S. Utz, M. Tanis & S. B. Barnes (Eds.), *Mediated Interpersonal Communication* (pp. 177-199). New York/London: Routledge.

Hartmann, T., & Goldhoorn, C. (2011). Horton and Wohl Revisited: Exploring Viewers' Experience of Parasocial Interaction. *Journal of Communication, 61*, 1104-1121.

Hausmann, R. G. M., & VanLehn, K. (2010). The effect of self-explaining on robust learning. *International Journal of Artificial Intelligence in Education, 20*(4), 303-332.

Heberlein, A. S., & Atkinson, A. P. (2009). Neuroscientific evidence for simulation and shared substrates in emotion recognition: Beyond faces. *Emotion Review, 1*(2), 162-177.

Hilbert, T. S., & Renkl, A. (2009). Learning how to use a computer-based concept-mapping tool: Self-explaining examples helps. *Computers in Human Behavior, 25*(2), 267-274. doi: http://dx.doi.org/10.1016/j.chb.2008.12.006

Hilbert, T. S., Renkl, A., Schworm, S., Kessler, S., & Reiss, K. (2008). Learning to teach with worked-out examples: A computer-based learning environment for teachers. *Journal of Computer Assisted Learning, 24*(4), 316-332.

Hodges, B., Regehr, G., & Martin, D. (2001). Difficulties in Recognizing One's Own Incompetence: Novice Physicians Who Are Unskilled and Unaware of It. *Academic Medicine, 76*(10), S87-S89.

Hogan, R. (1969). Development of an empathy scale. *Journal of Consulting and Clinical Psychology, 33*(3), 307-316. doi: 10.1037/h0027580

Homer, B. D., Plass, J. L., & Blake, L. (2008). The effects of video on cognitive load and social presence in multimedia-learning. *Computers in Human Behavior, 24*(3), 786-797. doi: 10.1016/j.chb.2007.02.009

Horton, D., & Wohl, R. R. (1956). Mass communication and para-social interaction. *Psychiatry, 19*, 215-229.

Howard-Jones, P. (2011). The impact of digital technologies on human wellbeing. Evidence from the science of mind and brain. A state of the art review (Publication no. http://www.nominettrust.org.uk/sites/default/files/SoAInternetandthebrain_0.pdf). Retrieved 01.09.2011, from Nominet Trust

Issenberg, S. B., McGaghie, W. C., Petrusa, E. R., Gordon, D. L., & Scalese, R. J. (2005). Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. *Medical Teacher, 27*(1), 10-28.

Kalyuga, S. (2011). Cognitive load theory: How many types of load does it really need? *Educational Psychology Review, 23*(1), 1-19. doi: 10.1007/s10648-010-9150-7

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 38*(1), 23-31. doi: 10.1207/S15326985EP3801_4

Kalyuga, S., Rikers, R., & Paas, F. (2012). Educational Implications of Expertise Reversal Effects in Learning and Performance of Complex Cognitive and Sensorimotor Skills. *Educational Psychology Review, 24*(2), 313-337. doi: 10.1007/s10648-012-9195-x

Kaufhold, M. (2006). *Kompetenz und Kompetenzerfassung. Analyse und Beurteilung von Verfahren der Kompetenzerfassung*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Khawaja, M. A., Chen, F., & Marcus, N. (2014). Measuring cognitive load using linguistic features: Implications for usability evaluation and adaptive interaction design. *International Journal of Human-Computer Interaction, 30*(5), 343-368. doi: 10.1080/10447318.2013.860579

Kim, C. (2012). The role of affective and motivational factors in designing personalized learning environments. *Educational Technology Research and Development, 60*(4), 563-584. doi: 10.1007/s11423-012-9253-6

Kim, C., Park, S. W., & Cozart, J. (2014). Affective and motivational factors of learning in online mathematics courses. *British Journal of Educational Technology, 45*(1), 171-185. doi: 10.1111/j.1467-8535.2012.01382.x

Klimmt, C., Hartmann, T., & Schramm, H. (2006). Parasocial interactions and relationships. In J. Bryant & P. Vorderer (Eds.), *Psychology of Entertainment* (pp. 291-313). Mahwah, NJ: Lawrence Earlbaum.

Kneebone, R., Nestel, D., Wetzel, C., Black, S., Jacklin, R., Aggarwal, R., . . . Darzi, A. (2006). The Human Face of Simulation: Patient-Focused Simulation Training. *Academic Medicine, 81*(10), 919-924. doi: 10.1097/01.ACM.0000238323.73623.c2

Kraiger, K., Ford, J. K., & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology, 78*(2), 311-328.

Kuhn, D., & Katz, J. (2009). Are self-explanations always beneficial? *Journal of Experimental Child Psychology, 103*(3), 386-394. doi: http://dx.doi.org/10.1016/j.jecp.2009.03.003

Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics, 33*(1), 159-174. doi: 10.2307/2529310

Lane, J. L., Slavin, S., & Ziv, A. (2001). Simulation in medical education: A review. *Simulation & Gaming, 32*(3), 297-314.

Laschinger, S., Medves, J., Pulling, C., McGraw, R., Waytuck, B., Harrison, M. B., & Gambeta, K. (2008). Effectiveness of simulation on health profession students' knowledge, skills, confidence and satisfaction. *International Journal of Evidence-Based Healthcare, 6*(3), 278-302.

Lee, E.-J., & Jang, J.-w. (2013). Not so imaginary interpersonal contact with public figures on social network sites: How affiliative tendency moderates its effects. *Communication Research, 40*(1), 27-51. doi: 10.1177/0093650211431579

Leibetseder, M., Laireiter, A.-R., & Köller, T. (2007). Structural analysis of the E-scale. *Personality and Individual Differences, 42*(3), 547-561. doi: 10.1016/j.paid.2006.08.002

Leibetseder, M., Laireiter, A.-R., Riepler, A., & Köller, T. (2001). E-Skala: Fragebogen zur Erfassung von Empathie - Beschreibung und psychometrische Eigenschaften. *Zeitschrift für Differentielle und Diagnostische Psychologie, 22*(1), 70-85.

Lenhard, W. (2009). Der Einsatz intelligenter tutorieller Systeme zur Vermittlung von Leseverständnisstrategien. In W. Lenhard & W. Schneider (Eds.), *Diagnose und Förderung von Leseverständnis und Lesekompetenz*. Göttingen: Hogrefe.

Leppink, J., Paas, F., Van der Vleuten, C. P. M., Van Gog, T., & Van Merriënboer, J. J. G. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods, 45*(4), 1058-1072.

Leppink, J., Paas, F., van Gog, T., van der Vleuten, C. P. M., & van Merriënboer, J. J. G. (2014). Effects of pairs of problems and examples on task performance and different types of cognitive load. *Learning and Instruction, 30*, 32-42. doi: 10.1016/j.learninstruc.2013.12.001

Lesho, E. D. O., Foster, L. D. O., Wang, Z. P., Sarmiento, D. M. D., Dennison, S. M. D., Vahey, M. T. P., . . . Smalls, C. R. N. (2009). The Accuracy of Physicians' Perceptions of Patients' Suffering: Findings From Two Teaching Hospitals. *Academic Medicine, 84*(5), 636-642.

Libert, Y., Conradt, S., Reynaert, C., Janne, P., Tordeurs, D., Delvaux, N., . . . Razavi, D. (2001). Improving doctor's communication skills in oncology: review and future perspectives. *Bulletin du cancer, 88*(12), 1167-1176.

Lim, J., Reiser, R., & Olina, Z. (2009). The effects of part-task and whole-task instructional approaches on acquisition and transfer of a complex cognitive skill. *Educational Technology Research & Development, 57*(1), 61-77. doi: 10.1007/s11423-007-9085-y

Loh, V., Andrews, S., Hesketh, B., & Griffin, B. (2013). The moderating effect of individual differences in error-management training: Who learns from mistakes? *Human Factors, 55*(2), 435-448. doi: 10.1177/0018720812451856

Lombard, M., & Ditton, T. (1997). At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication, 3*(2).

Lombard, M., Reich, R. D., Grabe, M. E., Bracken, C. C., & Ditton, T. B. (2000). Presence and Television. *Human Communication Research, 26*(1), 75.

Lombrozo, T. (2006). The structure and function of explanations. *Trends In Cognitive Sciences, 10*(10), 464-470.

Mayer, R. E. (2005). Principles of Multimedia Learning Based on Social Cues: Personalization, Voice, and Image Principles *The Cambridge handbook of multimedia learning.* (pp. 201-212). New York, NY US: Cambridge University Press.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 43-52. doi: 10.1207/S15326985EP3801_6

McKee, T. L. E., & Ptacek, J. T. (2001). I'm afraid I have something bad to tell you: Breaking bad news from the perspective of the giver. *Journal of Applied Social Psychology, 31*(2), 246-273.

McKenney, S., & Reeves, T. C. (2013). Systematic review of design-based research progress: Is a little knowledge a dangerous thing? *Educational Researcher, 42*(2), 97-100.

Mehrabian, A., & Epstein, N. (1972). A measure of emotional empathy. *Journal of Personality, 40*(4), 525-543. doi: 10.1111/1467-6494.ep8970028

Merker, B. M., Hanson, W. E., & Poston, J. M. (2010). National survey of psychologists' training and practice in breaking bad news: A mixed methods study of the MUM effect. *Journal of Clinical Psychology in Medical Settings, 17*(3), 211-219. doi: 10.1007/s10880-010-9197-0

Meunier, J., Merckaert, I., Libert, Y., Delvaux, N., Etienne, A.-M., Liénard, A., . . . Razavi, D. (2013). The effect of communication skills training on residents' physiological arousal in a breaking bad news simulated task. *Patient Education and Counseling, 93*(1), 40-47. doi: 10.1016/j.pec.2013.04.020

Op't Eynde, P., & Turner, J. E. (2006). Focusing on the complexity of emotion issues in academic learning: A dynamical component systems approach. *Educational Psychology Review, 18*(4), 361-376. doi: 10.1007/s10648-006-9031-2

Opfermann, M. (2008). *There's more to it than instructional design: The role of individual learner characteristics for hypermedia learning.* IWM KMRC, Logos Berlin.

Paas, F. G. W. C. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology, 84*(4), 429-434.

Paas, F. G. W. C., & Van Merriënboer, J. J. G. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review, 6*(4), 351-371.

Paas, F. G. W. C., Van Merriënboer, J. J. G., & Adam, J. J. (1994). Measurement of cognitive load in instructional research. *Perceptual and Motor Skills, 79*(1), 419-430. doi: 10.2466/pms.1994.79.1.419

Parle, M., Maguire, P., & Heaven, C. (1997). The development of a training model to improve health professionals' skills, self-efficacy and outcome expectancies when communicating with cancer patients. *Social Science & Medicine, 44*(2), 231-240.

Paskins, Z., & Peile, E. (2010). Final year medical students' views on simulation-based teaching: A comparison with the Best Evidence Medical Education Systematic Review. *Medical Teacher, 32*, 569-577.

Patrício, M. F., Julião, M., Fareleira, F., & Carneiro, A. V. (2013). Is the OSCE a feasible tool to assess competencies in undergraduate medical education? *Medical Teacher, 35*(6), 503-514. doi: doi:10.3109/0142159X.2013.774330

Pekrun, R., Götz, T., & Perry, R. P. (2005). Academic Emotions Questionnaire (AEQ).

Pekrun, R., Götz, T., Perry, R. P., Kramer, K., & Hochstadt, M. (2004). Beyond test anxiety: Development and validation of the Test Emotions Questionnaire (TEQ). *Anxiety, Stress & Coping, 17*, 287-316.

Pieters, H., Touw-Otten, F., & De Melker, R. (1994). Simulated patients in assessing consultation skills of trainees in general practice vocational training. *Medical Education, 28*, 226-233.

Pirolli, P., & Recker, M. (1994). Learning Strategies and Transfer in the Domain of Programming. *Cognition and Instruction, 12*(3), 235-275. doi: 10.1207/s1532690xci1203_2

Preston, S. D., & de Waal, F. B. M. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences, 25*, 1-72.

Ptacek, J. T., & Eberhardt, T. L. (1996). Breaking bad news. A review of the literature. *Jama, 276*(6), 496-502.

Ptacek, J. T., & McIntosh, E. G. (2009). Physician challenges in communicating bad news. *Journal of Behavioral Medicine, 32*(4), 380-387.

Ptacek, J. T., Ptacek, J. J., & Ellison, N. M. (2001). 'I'm sorry to tell you ...' Physicians' reports of breaking bad news. *Journal of Behavioral Medicine, 24*(2), 205-217.

Radey, M., & Figley, C. R. (2007). The social psychology of compassion. *Clinical Social Work Journal, 35*(3), 207-214.

Realdon, O., Zurloni, V., Confalonieri, L., Mantovani, F., Riva, G., Anguera, M. T., . . . Mantovani, F. (2006). Learning communication skills through computer-based interactive simulations *From communication to presence: Cognition, emotions and culture towards the ultimate communicative experience: Festschrift in honor of Luigi Anolli.* (pp. 276-298). Amsterdam Netherlands: IOS Press.

Reisslein, J., Sullivan, H., & Reisslein, M. (2007). Learner Achievement and Attitudes under Different Paces of Transitioning to Independent Problem Solving. *Journal of Engineering Education, 96*(1), 45-55.

Renkl, A. (1997). Learning from worked-out examples: A study on individual differences. *Cognitive Science, 21*, 1-29.

Renkl, A. (2005). The worked-out examples principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 229-245). Cambridge: Cambridge University Press.

Renkl, A. (2014). Toward an instructionally oriented theory of example-based learning. *Cognitive Science, 38*(1), 1-37. doi: 10.1111/cogs.12086

Renkl, A., & Atkinson, R. K. (2003). Structuring the transition from example study to problem solving in cognitive skill acquisition: A cognitive load perspective. *Educational Psychologist, 38*(1), 15-22.

Renkl, A., Atkinson, R. K., Maier, U. H., & Staley, R. (2002). From example study to problem solving: smooth transitions help learning. *The journal of Experimental Education, 70*(4), 293-315.

Renkl, A., Mandl, H., & Gruber, H. (1996). Inert knowledge: Analyses and remedies. *Educational Psychologist, 31*(2), 115-121. doi: 10.1207/s15326985ep3102_3

Renkl, A., Stark, R., Gruber, H., & Mandl, H. (1998). Learning from worked-out examples: The effects of example variability and elicited self-explanations. *Contemporary Educational Psychology, 23*(1), 90-108. doi: 10.1006/ceps.1997.0959

Rittle-Johnson, B., Star, J. R., & Durkin, K. (2009). The importance of prior knowledge when comparing examples: Influences on conceptual and procedural knowledge of equation solving. *Journal of Educational Psychology, 101*(4), 836-852. doi: 10.1037/a0016026

Rosen, S., & Tesser, A. (1970). On reluctance to communicate undesirable information: The MUM effect. *Sociometry, 33*(3), 253-263. doi: 10.2307/2786156

Rosenzweig, S., Reibel, D. K., Greeson, J. M., Brainard, G. C., & Hojat, M. (2003). Mindfulness-Based Stress Reduction Lowers Psychological Distress In Medical Students. *Teaching & Learning in Medicine, 15*(2), 88-92.

Rourke, A., & Sweller, J. (2009). The worked-example effect using ill-defined problems: Learning to recognise designers' styles. *Learning and Instruction, 19*(2), 185-199. doi: http://dx.doi.org/10.1016/j.learninstruc.2008.03.006

Roy, M., & Chi, M. T. H. (2005). The Self-Explanation Principle in Multimedia Learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning.* (pp. 271-286). New York, NY US: Cambridge University Press.

Rubin, A. M., Perse, E. M., & Powell, R. A. (1985). Loneliness, Parasocial Interaction and Local Television Viewing. *Human Communication Research, 12*, 155-180.

Rummel, N., & Spada, H. (2005). Learning to collaborate: An instructional approach to promoting collaborative problem solving in computer-mediated settings. [Lernen kooperativ zu arbeiten: Ein Unterrichtsansatz zur Förderung kooperativen Problemlösens in computervermittelten Settings]. *Journal of the Learning Sciences, 14*(2), 201-241.

Rummel, N., Spada, H., & Hauser, S. (2009). Learning to collaborate while being scripted or by observing a model. *International Journal of Computer-Supported Collaborative Learning, 4*, 69-92.

Salas, E., Rosen, M. A., Held, J. D., & Weissmuller, J. J. (2009). Performance measurement in simulation-based training: A review and best practices. *Simulation & Gaming, 40*(3), 328-376.

Salden, R. J. C. M., Koedinger, K. R., Renkl, A., Aleven, V., & McLaren, B. M. (2010). Accounting for Beneficial Effects of Worked Examples in Tutored Problem Solving. *Educational Psychology Review, 22*(4), 379-392. doi: 10.1007/s10648-010-9143-6

Salomon, G. (1981). Communication and education: Social and psychological interactions. *People & Communication, 13*, 9-271.

Salomon, G. (1992). New challenges for educational research: Studying the individual within learning environments. *Scandinavian Journal of Educational Research, 36*(3), 167-182.

Salvucci, D. D. (2013). Integration and Reuse in Cognitive Skill Acquisition. *Cognitive Science, 37*(5), 829-860. doi: 10.1111/cogs.12032

Sarason, I. G., Peterson, A. M., & Nyman, B. (1968). Test anxiety and the observation of models. *Journal of Personality, 36*(3), 493-511. doi: 10.1111/j.1467-6494.1968.tb01488.x

Scardamalia, M. (2001). Big Change Questions Will Educational Institutions, within their Present Structures, be Able to Adapt Sufficiently to Meet the Needs of the Information Age? *Journal of Educational Change, 2*(2), 171-176. doi: 10.1023/A:1017982102060

Schachter, S., Singer, J. E., & Parrott, W. G. (2001). Cognitive, social, and psychological determinants of emotional state *Emotions in social psychology: Essential readings.* (pp. 76-93). New York, NY US: Psychology Press.

Schank, R. C. (Ed.). (1997). *Virtual Learning: A revolutionary approach to building a highly skilled workforce*. New York: McGraw-Hill.

Scheiter, K., Gerjets, P., Huk, T., Imhof, B., & Kammerer, Y. (2009). The effects of realism in learning with dynamic visualizations. *Learning and Instruction, 19*(6), 481-494.

Scherer, K. R. (2000). Psychological models of emotion. In J. Borod (Ed.), *The neuropsychology of emotion* (pp. 137-162). Oxford, England and New York: Oxford University Press.

Scherer, Y. K., Bruce, S. A., & Runkawatt, V. (2007). A Comparison of Clinical Simulation and Case Study Presentation on Nurse Practitioner Students' Knowledge and Confidence in Managing a Cardiac Event. *International Journal of Nursing Education Scholarship, 4*(1), 1-14.

Schiappa, E., Gregg, P. B., & Hewes, D. E. (2005). The Parasocial Contact Hypothesis. *Communication Monographs, 72*(1), 92-115.

Schiappa, E., Gregg, P. B., & Hewes, D. E. (2006). Can One TV Show Make a Difference?a Will & Grace and the Parasocial Contact Hypothesis. *Journal of Homosexuality, 51*(4), 15-37.

Schramm, H., & Hartmann, T. (2008). The PSI-Process Scales. A new measure to assess the intensity and breadth of parasocial processes. *Communications, 33*(4), 385-401.

Schramm, H., & Wirth, W. (2010). Testing a universal tool for measuring parasocial interactions across different situations and media: Findings from three studies. *Journal of Media Psychology: Theories, Methods, and Applications, 22*(1), 26-36.

Schunk, D. H. (1996). Goal and self-evaluative influences during children's cognitive skill learning. *American Educational Research Journal, 33*(2), 359-382.

Schwamborn, A., Thillmann, H., Opfermann, M., & Leutner, D. (2011). Cognitive load and instructionally supported learning with provided and learner-generated visualizations. *Computers in Human Behavior, 27*(1), 89-93. doi: 10.1016/j.chb.2010.05.028

Schwonke, R., Renkl, A., Salden, R., & Aleven, V. (2011). Effects of different ratios of worked solution steps and problem solving opportunities on cognitive load and learning outcomes. *Computers in Human Behavior, 27*(1), 58-62. doi: 10.1016/j.chb.2010.03.037

Schworm, S., & Renkl, A. (2006). Computer-Supported Example-Based Learning: When Instructional Explanations Reduce Self-Explanations. *Computers and Education, 46*(4), 426-445.

Schworm, S., & Renkl, A. (2007). Learning argumentation skills through the use of prompts for self-explaining examples. *Journal of Educational Psychology, 99*(2), 285-296.

Seidel, T., Blomberg, G., & Renkl, A. (2013). Instructional strategies for using video in teacher education. *Teaching & Teacher Education, 34*, 56-65. doi: 10.1016/j.tate.2013.03.004

Shen, C.-Y., & Tsai, H.-C. (2009). Design Principles of Worked Examples: A Review of the Empirical Studies. *Journal of Instructional Psychology, 36*(3), 238-244.

Siebeck, M., Schwald, B., Frey, C., Roding, S., Stegmann, K., & Fischer, F. (2011). Teaching the rectal examination with simulations: effects on knowledge acquisition and inhibition. *Med Educ, 45*(10), 1025-1031. doi: 10.1111/j.1365-2923.2011.04005.x

Siegler, R. S. (1995). How Does Change Occur: A Microgenetic Study of Number Conservation. *Cognitive Psychology, 28*(3), 225-273. doi: http://dx.doi.org/10.1006/cogp.1995.1006

Siegler, R. S. (2002). Microgenetic studies of self-explanation. In N. Granott & J. Parziale (Eds.), *Microdevelopment: Transition processes in development and learning.* (pp. 31-58). New York, NY US: Cambridge University Press.

Spiro, R. J., Collins, B. P., & Ramchandran, A. (2007). Reflections on a post-Gutenberg epistemology for video use in ill-structured domains: Fostering complex learning and cognitive flexibility. In R. Goldman, R. D. Pea, B. Barron & S. Derry (Eds.), *Video research in the learning sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.

Stark, R. (1999). *Lernen mit Lösungsbeispielen. Einfluss unvollständiger Lösungsbeispiele auf Beispielelaboration,*
Lernerfolg und Motivation. Göttingen: Hogrefe.

Stark, R. (2000). Instruktionale Effekte beim Lernen mit unvollständigen Lösungsbeispielen *Forschungsbericht*. München: Ludwig-Maximilians Universität, Lehrstuhl für Empirische Pädagogik und Pädagogische Psychologie.

Stark, R., Gruber, H., Renkl, A., & Mandl, H. (2000). Instruktionale Effekte einer kombinierten Lernmethode: Zahlt sich die Kombination von Lösungsbeispielen und

Problemlöseaufgaben aus? *Zeitschrift für Pädagogische Psychologie/ German Journal of Educational Psychology, 14*(4), 206-218.

Stark, R., Kopp, V., & Fischer, M. R. (2011). Case-based learning with worked examples in complex domains: Two experimental studies in undergraduate medical education. *Learning and Instruction, 21*(1), 22-33. doi: 10.1016/j.learninstruc.2009.10.001

Stark, R., Mandl, H., Gruber, H., & Renkl, A. (1999). Instructional means to overcome transfer problems in the domain of economics: Empirical studies. *International Journal of Educational Research, 31*(7), 591-609.

Steele-Johnson, D., & Kalinoski, Z. T. (2014). Error framing effects on performance: Cognitive, motivational, and affective pathways. *The Journal of Psychology: Interdisciplinary and Applied, 148*(1), 93-111. doi: 10.1080/00223980.2012.748581

Stenning, K., McKendree, J., Lee, J., Cox, R., Dineen, F., & Mayes, T. (1999). *Vicarious Learning from Educational Dialogue.* Paper presented at the conference on Computer support for collaborative learning CSCL.

Stotland, E. (1969). Exploratory investigations of empathy. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 4). New York: Academic Press.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science: A Multidisciplinary Journal, 12*(2), 257-285. doi: 10.1207/s15516709cog1202_4

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*(4), 295-312. doi: 10.1016/0959-4752(94)90003-5

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review, 22*(2), 123-138.

Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction, 2*(1), 59-89. doi: 10.1207/s1532690xci0201_3

Sweller, J., Van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251-296. doi: 10.1023/A:1022193728205

Taatgen, N. A. (2013). The Nature and Transfer of Cognitive Skill. *Psychological Review, 120*(3), 439-471. doi: 10.1037/a0033138

Tesser, A., & Rosen, S. (1975). The reluctance to transmit bad news. In L. Berkowitz (Ed.), *Advances in experimental social psychology, Vol. 8.* San Diego, CA US: Academic Press.

Tesser, A., Rosen, S., & Batchelor, T. R. (1972). On the reluctance to communicate bad news (the mum effect): A role play extension. *Journal of Personality, 40*(1), 88-103.

Tesser, A., Rosen, S., & Tesser, M. (1971). On the reluctance to communicate undesirable messages (the MUM effect): A field study. *Psychological Reports, 29*(2), 651-654.

Titchener, E. B. (1909). Lectures on the Experimental Psychology of the Thought-processes (pp. 318): University of Michigan.

Turner, E. V., Helper, M. M., & Kriska, S. D. (1974). Predictors of clinical performance. *Journal of Medical Education, 49*(4), 338-342.

van Gog, T., Kester, L., & Paas, F. (2011). Effects of worked examples, example-problem, and problem-example pairs on novices' learning. *Contemporary Educational Psychology, 36*(3), 212-218. doi: 10.1016/j.cedpsych.2010.10.004

van Gog, T., Paas, F., & Van Merriënboer, J. J. G. (2004). Process-Oriented Worked Examples: Improving Transfer Performance through Enhanced Understanding. *Instructional Science: An International Journal of Learning and Cognition, 32*(1-2), 83-98.

van Gog, T., & Rummel, N. (2010). Example-Based Learning: Integrating Cognitive and Social Cognitive Research Perspectives. *Educational Psychology Review, 22*, 155-174.

Van Merriënboer, J. J. G., Clark, R. E., & de Croock, M. B. M. (2002). Blueprints fo complex learning: The 4C/ID-Model. *Educational Technology Research and Development, 50*(2), 39-64. doi: 10.1007/BF02504993

Van Merriënboer, J. J. G., & de Croock, M. B. (1992). Strategies for computer-based programming instruction: Program completion vs. program generation. *Journal of Educational Computing Research, 8*(3), 365-394. doi: 10.2190/MJDX-9PP4-KFMT-09PM

Van Meter, P. (2001). Drawing construction as a strategy for learning from text. *Journal of Educational Psychology, 93*(1), 129-140. doi: 10.1037/0022-0663.93.1.129

van Ophuysen, S., & Hannover, B. (2005). Die Wirkung von Modellpersonen auf das bereichsspezifische Selbstwissen des Rezipienten. *Zeitschrift für Psychologie, 213*(1), 1-8.

van Zundert, M. J., Sluijsmans, D. M. A., Könings, K. D., & Van Merriënboer, J. J. G. (2012). The differential effects of task complexity on domain-specific and peer assessment skills. *Educational Psychology, 32*(1), 127-145. doi: 10.1080/01443410.2011.626122

VanLehn, K. (1996). Cognitive skill acquisition. *Annual Review of Psychology, 47*(1), 513.

VanLehn, K. (1999). Rule-Learning Events in the Acquisition of a Complex Skill: An Evaluation of Cascade. *Journal of the Learning Sciences, 8*(1), 71.

VanLehn, K. (2006). The behavior of tutoring systems. *International Journal of Artificial Intelligence in Education, 16*(3), 227-265.

VanLehn, K., Jones, R. M., & Chi, M. T. H. (1992). A Model of the Self-Explanation Effect. *Journal of the Learning Sciences, 2*(1), 1.

Wang, F., & Hannafin, M. (2005). Design-based research and technology-enhanced learning environments. *Educational Technology Research and Development, 53*(4), 5-23. doi: 10.1007/BF02504682

Wild, B., Erb, M., & Bartels, M. (2001). Are emotions contagious? Evoked emotions while viewing emotionally expressive faces: quality, quantity, time course and gender differences. *Psychatrie Research, 102*, 109-124.

Wood, D. (2010). Simulation training: Preparing clinical teams to improve patient care. *Telemedicine and e-Health, 16*(4), 400-404.

Yeh, Y.-F., Chen, M.-C., Hung, P.-H., & Hwang, G.-J. (2010). Optimal self-explanation prompt design in dynamic multi-representational learning environments. *Computers & Education, 54*(4), 1089-1100.

Zeidner, M. (1998). *Test anxiety: The state of the art*. New York, NY US: Plenum Press.

Zeidner, M. (2007). Test anxiety in educational contexts: Concepts, findings, and future directions. In P. A. Schutz & R. Pekrun (Eds.), *Emotion in education.* (pp. 165-184). San Diego, CA US: Elsevier Academic Press.

Zendejas, B., Brydges, R., Wang, A. T., & Cook, D. A. (2013). Patient outcomes in simulation-based medical education: A systematic review. *Journal of General Internal Medicine, 28*(8), 1078-1089. doi: 10.1007/s11606-012-2264-5

Zillmann, D. (2006). Empathy: Affective Reactivity to Others' Emotional Experiences. In J. Bryant & P. Vorderer (Eds.), *Psychology of Entertainment* (pp. 151-181). Mahwah, NJ: Lawrence Earlbaum.

Zillmann, D., & Cantor, J. R. (1977). Affective responses to the emotions of a protagonist. *Journal of Experimental Social Psychology, 13*(2), 155-165. doi: http://dx.doi.org/10.1016/S0022-1031(77)80008-5

List of Figures

List of Tables

Appendices

Appendix 1: Plan for SP week with introduction to breaking bad news.

MeCuM LMU

Modul 3

7M1201 **Plan für die SP-Woche Chirurgie Innenstadt und Chirurgie Großhadern**

	Mo	Die	Mi	Do	Fr
8:00 – 13:00	**8:30 – 11:30 h** **~~Rektale Untersuchung am Modell~~ und Einführung in die Übung „schwieriges Gespräch"** **~~Siehe Moodle und~~ Hörsaalaushänge**			**8:00 – 12:30** **Rektale Untersuchung am Stand Pat. Ambulanz Nussbaumstr. Raum 05**	**9:00 – 11:30 h** **Praktische Prüfung Aufklärungsgespräch Chirurgie Innenstadt, Seminarräume 4/5**
13:00-13:45	Hauptvorlesung	Hauptvorlesung		Hauptvorlesung	Hauptvorlesung
14:30 – 18:30		**Übung schwieriges Gespräch mit Stand.**		**14:30 – 18:30** **Rektale Untersuchung am Stand Pat. Ambulanz Nussbaumstr. Raum 05**	
17:15-18:00	Hauptvorlesung				

Appendix 2: Short version of the spikes text based on Baile and colleagues (2000).

SPIKES—A Six-Step Protocol for Delivering Bad News

<u>A definition of bad news</u>

Bad news may be defined as "any information which adversely and seriously affects an individual's view of his or her future". **Bad news is always, however, in the "eye of the beholder,"** such that one cannot estimate the impact of the bad news until one has first determined the recipient's expectations or understanding. For example, a patient who is told that her back pain is caused by a recurrence of her breast cancer when she was expecting to be told it was a muscle strain is likely to feel shocked.

<u>It is important because it is a frequent but stressful task.</u>

Over the course of a career, a busy clinician may disclose unfavourable medical information to patients and families many thousands of times. Breaking bad news to cancer patients is inherently aversive, described as "hitting the patient over the head" or "dropping a bomb". Breaking bad news can be particularly stressful when the clinician is inexperienced, the patient is young, or there are limited prospects for successful treatment.

The process of disclosing bad news can be viewed as an attempt to achieve four essential goals.

1. Gathering information from the patient. This allows the physician to determine the patient's knowledge and expectations and readiness to hear the bad news.
2. Provide intelligible information in accordance with the patient's needs and desires.
3. Support the patient by employing skills to reduce the emotional impact and isolation experienced by the recipient of bad news.
4. Develop a strategy in the form of a treatment plan with the input and cooperation of the patient.

Not every episode of breaking bad news will require all of the steps of SPIKES, but when they do they are meant to follow each other in sequence.

THE SIX STEPS OF SPIKES

STEP 1: S—SETTING UP the Interview

Mental rehearsal is a useful way for preparing for stressful tasks. It is helpful to be reminded that, although bad news may be very sad for the patients, the information may be important in allowing them to plan for the future.

Some helpful guidelines:

- *Arrange for some privacy*.
  An interview room is ideal, but, if one is not available, draw the curtains around the patient's bed. Have tissues ready in case the patient becomes upset.
- *Involve significant others*.
  Most patients want to have someone else with them but this should be the patient's choice. When there are many family members, ask the patient to choose one or two family representatives.
- *Sit down*.
  Sitting down relaxes the patient and is also a sign that you will not rush. When you sit, try not to have barriers between you and the patient. If you have recently examined the patient, allow them to dress before the discussion.
- *Make connection with the patient*.
  Maintaining eye contact may be uncomfortable but it is an important way of establishing rapport. Touching the patient on the arm or holding a hand (if the patient is comfortable with this) is another way to accomplish this.
- *Manage time constraints and interruptions*.

Inform the patient of any time constraints you may have or interruptions you expect. Set your pager on silent or ask a colleague to respond to your pages.

STEP 2: P—ASSESSING THE PATIENT'S PERCEPTION
Steps 2 and 3 of SPIKES are points in the interview where you implement the axiom "before you tell, ask." That is, before discussing the medical findings, the clinician uses open-ended questions to create a reasonably accurate picture of how the patient perceives the medical situation— what it is and whether it is serious or not.

For example:
"What have you been told about your medical situation so far?" or "What is your understanding of the reasons we did the MRI?".

Based on this information you can correct misinformation and tailor the bad news to what the patient understands. It can also accomplish the important task of determining if the patient is engaging in any variation of illness denial: wishful thinking, omission of essential but unfavourable medical details of the illness, or unrealistic expectations of treatment.

STEP 3: I—OBTAINING THE PATIENT'S INVITATION
While a majority of patients express a desire for full information about their diagnosis, prognosis, and details of their illness, some patients do not.
When a clinician hears a patient express explicitly a desire for information, it may lessen the anxiety associated with divulging the bad news. However, shunning information is a valid psychological coping mechanism and may be more likely to be manifested as the illness becomes more severe.

For example:
"How would you like me to give the information about the test results? Would you like me to give you all the information or sketch out the results and spend more time discussing the treatment plan?".

If patients do not want to know details, offer to answer any questions they may have in the future or to talk to a relative or friend.

STEP 4: K—GIVING KNOWLEDGE AND INFORMATION TO THE PATIENT
Warning the patient that bad news is coming may lessen the shock that can follow the disclosure of bad news and may facilitate information processing.

For example:
"Unfortunately I've got some bad news to tell you" or "I'm sorry to tell you that…".

Simple guidelines for giving medical facts:

1. Start at the level of comprehension and vocabulary of the patient.

2. Try to use nontechnical words such as "spread" instead of "metastasized" and "sample of tissue" instead of "biopsy."

3. Avoid excessive bluntness (e.g., "You have very bad cancer and unless you get treatment immediately you are going to die.") as it is likely to leave the patient isolated and later angry, with a tendency to blame the messenger of the bad news.

4. Give information in small chunks and check periodically as to the patient's understanding.

5. When the prognosis is poor, avoid using phrases such as “There is nothing more we can do for you.” This attitude is inconsistent with the fact that patients often have other important therapeutic goals such as good pain control and symptom relief.

STEP 5: E—ADDRESSING THE PATIENT’S EMOTIONS WITH EMPATHIC RESPONSES

Responding to the patient’s emotions is one of the most difficult challenges of breaking bad news. Patients’ emotional reactions may vary from silence to disbelief, crying, denial, or anger. When patients get bad news their emotional reaction is often an expression of shock, isolation, and grief. In this situation the physician can offer support and solidarity to the patient by making an empathic response.

An empathic response consists of four steps:

1. Observe for any emotion on the part of the patient. This may be tearfulness, a look of sadness, silence, or shock.

2. Identify the emotion experienced by the patient by naming it to oneself. If a patient appears sad but is silent, use open questions to query the patient as to what they are thinking or feeling.

3. Identify the reason for the emotion. This is usually connected to the bad news. However, if you are not sure, again, ask the patient.

4. After you have given the patient a brief period of time to express his or her feelings, let the patient know that you have connected the emotion with the reason for the emotion by making a connecting statement.

An example:

Doctor: I’m sorry to say that the x-ray shows that the chemotherapy doesn’t seem to be working [pause]. Unfortunately, the tumour has grown somewhat.

Patient: I’ve been afraid of this! [Cries]

Doctor: [Moves his chair closer, offers the patient a tissue, and pauses.] I know that this isn’t what you wanted to hear. I wish the news were better.

In the above dialogue, the physician observed the patient crying and realized that the patient was tearful because of the bad news. He moved closer to the patient. At this point he might have also touched the patient’s arm or hand if they were both comfortable and paused a moment to allow her to get her composure. He let the patient know that he understood why she was upset by making a statement that reflected his understanding.

Table 2. Examples of empathic, exploratory, and validating responses

Empathic statements	Exploratory questions	Validating responses
“I can see how upsetting this is to you.”	“How do you mean?”	“I can understand how you felt that way.”
“I can tell you weren’t expecting to hear this.”	“Tell me more about it.”	“I guess anyone might have that same reaction.”
“I know this is not good news for you.”	“Could you explain what you mean?”	“You were perfectly correct to think that way.”
“I’m sorry to have to tell you this.”	“You said it frightened you?”	“Yes, your understanding of the reason for the tests is very good.”
“This is very difficult for me also.”	“Could you tell me what you’re worried about?”	“It appears that you’ve thought things through very well.”
“I was also hoping for a better result.”	“Now, you said you were concerned about your children. Tell me more.”	“Many other patients have had a similar experience.”

Until an emotion is cleared, it will be difficult to go on to discuss other issues. If the emotion does not diminish shortly, it is helpful to continue to make empathic responses (cf. Table 2) until the patient becomes calm.

Clinicians can also use empathic responses to acknowledge their own sadness or other emotions ("I also wish the news were better"). It can be a show of support to follow the empathic response with a validating statement, which lets the patient know that their feelings are legitimate.

Again, when emotions are not clearly expressed, such as when the patient is silent, the physician should ask an *exploratory question* before he makes an *empathic response* (see Table 2). When emotions are subtle or indirectly expressed or disguised as in thinly veiled disappointment or anger ("I guess this means I'll have to suffer through chemotherapy again") you can still use an *empathic response* ("I can see that this is upsetting news for you").

Patients regard their oncologist as one of their most important sources of psychological support, and combining empathic, exploratory, and validating statements is one of the most powerful ways of providing that support (Table 2). It reduces the patient's isolation, expresses solidarity, and validates the patient's feelings or thoughts as normal and to be expected.

STEP 6: S—STRATEGY AND SUMMARY

Patients who have a clear plan for the future are less likely to feel anxious and uncertain. Before discussing a treatment plan, it is important to ask patients if they are ready at that time for such a discussion. Presenting treatment options to patients when they are available is not only a legal mandate in some cases, but it will establish the perception that the physician regards their wishes as important. Sharing responsibility for decision-making with the patient may also reduce any sense of failure on the part of the physician when treatment is not successful. Checking the patient's misunderstanding of the discussion can prevent the documented tendency of patients to overestimate the efficacy or misunderstand the purpose of treatment. Clinicians are often very uncomfortable when they must discuss prognosis and treatment options with the patient, if the information is unfavourable.

Strategies to facilitate difficult discussions:

1. Many patients already have some idea of the seriousness of their illness and of the limitations of treatment but are afraid to bring it up or ask about outcomes. Exploring the patient's knowledge, expectations, and hopes (step 2 of SPIKES) will allow the physician to understand where the patient is and to start the discussion from that point. When patients have unrealistic expectations (e.g., "They told me that you work miracles."), asking the patient to describe the history of the illness will usually reveal fears, concerns, and emotions that lie behind the expectation. Patients may see cure as a global solution to several different problems that are significant for them. These may include loss of a job, inability to care for the family, pain and suffering, hardship on others, or impaired mobility. Expressing these fears and concerns will often allow the patient to acknowledge the seriousness of their condition. If patients become emotionally upset in discussing their concerns, it would be appropriate to use the strategies outlined in step 5 of SPIKES.

2. Understanding the important specific goals that many patients have, such as symptom control, and making sure that they receive the best possible treatment and continuity of care will allow the physician to frame hope in terms of what it is possible to accomplish. This can be very reassuring to patients.

Appendix 3: Briefing for the simulation.

Szenario Breaking Bad News

Sie sind Assistenzarzt im 2. Jahr in der Ambulanz. Der Patient wurde letzte Woche von Ihrem Kollegen, der jetzt im Urlaub ist, untersucht und anschließend zur Coloskopie geschickt. Ihr Kollege hat den Patienten für heute zu Ihnen zur Besprechung der Befunde wieder einbestellt.

Der Coloskopiebefund lautet:
Ileum und Valvula Bauhini unauffällig. Polypöser Tumor, 1/3 der Zirkumferenz umfassend, ca. 30 cm ab ano. Biopsie entnommen. Hämorrhoiden Grad 2. Der histologische Befund lautet wie folgt:
Mäßig differenziertes Adenokarzinom.

Ihre Aufgabe ist es nun, dem Patienten die Diagnose mitzuteilen und das weitere Vorgehen zu besprechen.

Medizinische Hintergrundinformation

Sie haben bereits den histologischen Befund des Patienten (mäßig differenziertes Adenokarzinom) erhalten und möchten sich zuvor die wesentlichen Schritte des weiteren Vorgehens vergegenwärtigen.

Staging: In einem ersten Schritt können eine abdominelle Sonografie, um evtl. verdächtige Lymphknoten zu erfassen, oder eine Spiral-CT durchgeführt werden. In einem nächsten Schritt ist evtl. eine Cystoskopie oder eine gynäkologische Untersuchung bei V.a. auf eine Infiltration der weiblichen Geschlechtsorgane notwendig.

Therapie: Die geplante Therapie richtet sich nach den Ergebnissen des Stagings.

Operation
Das übliche Vorgehen setzt sich aus der (radikalen) Sigmaresektion sowie der Entfernung des regionalen Lymphabflussgebietes zusammen. Ziel der Operation ist die vollständige und großzügige Entfernung des befallenen Darmteils. Sollten Nachbarorgane durch den Tumor erfasst sein, müssen diese ebenfalls mit dem befallenen Darmabschnitt en bloc reseziert werden. Bei schwierigen operativen Bedingungen kann es notwendig sein vorübergehend einen Anus praeter anzulegen. Falls der Tumor nicht operiert wird, besteht vor allem die Gefahr eines Ileus, einer Perforation oder von Symptomen der erfassten Nachbarorgane, da zu erwarten ist, dass der Tumor weiter wächst.

Adjuvante Chemo- und Radiotherapie
„Voraussetzung für eine adjuvante Therapie ist die R0-Resektion des Primärtumors und seines Lymphabflussgebietes. ... Für Patienten mit einem Kolonkarzinom im Stadium I und II oder nach R0-Resektion von Fernmetastasen ist eine adjuvante Therapie außerhalb von Studien nicht indiziert“.
Die postoperative Radio-/Chemotherapie verbessert im Stadium III die 5-Jahresüberlebensrate um ca. 10% und reduziert die Entstehung von Lokalrezidiven um ca. 50%.

Langzeit-Nebenwirkungen der Bestrahlung: Stenosen, Schrumpfblase, Fisteln, chronische Proktitis

Prognose:
abhängig vom Staging:

UICC- Stadium	**5-Jahresüberlebensrate**
0 Carcinoma in situ	
I max. bis zur Infiltration der Muscularis propria	**bis 95%**
II Infiltration aller Wandschichten bis zur Überschreitung der Darmwand	**bis 90%**
III Regionale Lymphknoten oder Infiltration der Umgebung	**bis 65%**
IV Fernmetastasen	**bis 5%**

Appendix 4: Factual knowledge test including four questions.

1. Bitte erläutern Sie, wie man sich im Schritt "S—SETTING UP the Interview" verhalten soll. **Welche Teilaufgaben beinhaltet dieser Schritt?**
2. Bitte erläutern Sie, wie man sich im Schritt "P—ASSESSING THE PATIENT'S PERCEPTION" verhalten soll. **Welche Teilaufgaben beinhaltet dieser Schritt?**

3. Bitte erläutern Sie, wie man sich im Schritt "E—ADDRESSING THE PATIENT'S EMOTIONS WITH EMPATHIC RESPONSES" verhalten soll. **Welche Teilaufgaben beinhaltet dieser Schritt?**
4. Bitte geben Sie drei Beispiele für emphatische Aussagen!

Appendix 5: Application-Oriented Knowledge test: Formulation of the questions following a short video clip.

	SPIKES Step	Task formulation
1.	P	Die Phase der Begrüßung ist abgeschlossen. Wie sollte der Arzt sich jetzt verhalten?
2.	E	Wie sollte der Arzt sich jetzt verhalten?
3.	Z	Das Gespräch nähert sich dem Ende. Wie sollte der Arzt sich jetzt verhalten?

Appendix 6: Coding instructions for the performance of breaking bad news based on the SPIKES by Baile and colleagues (2000).

S	**S1 Private Atmosphäre herstellen (arrange for privacy)** Trifft zu: Der Arzt bemüht sich darum eine private Atmosphäre herzustellen zum Beispiel indem er die Tür schließt und Taschentücher auf dem Tisch bereit stellt falls der Patient diese benötigt. Trifft nicht zu: Der Arzt lässt die Tür offen, verhindert nicht, dass andere Personen den Raum betreten und halt keine Taschentücher für den Patienten bereit.
	S2 Hinsetzen (sit down) *generell kodieren!* Trifft zu: Der Arzt zeigt dem Patienten wo er sitzen kann indem er auf einen Stuhl zeigt und sich schließlich auch selbst hinsetzt. Außerdem kann der Arzt die Stühle so anordnen, dass er und der Patient sich gegenüber sitzen und nicht zu weit voneinander entfernt sind. Trifft nicht zu: Der Arzt bleibt stehen oder vermeidet offensichtlich dem Patienten zu nah zu kommen indem er Objekte zwischen sie platziert (z.B. Wasserflasche oder Unterlagen) oder weit vom Patienten entfernt sitzt.
	S3 Anbieten, wichtige Andere einzubeziehen (Offer choice to involve significant others) Trifft zu: Der Arzt versucht herauszufinden ob der Patient alleine oder in Begleitung zu dem Termin gekommen ist. Falls er Jemanden mitgebracht hat, fragt der Arzt, ob der Patient diese Person gerne dabei hätte und lädt entsprechend ins Zimmer ein. Der Arzt muss anbieten einen Angehörigen dazu zu holen oder auch anbieten einen gesonderten Termin zu vereinbaren, bei dem Angehörige dabei sein können. Es ist egal an welcher Stelle im Gespräch nachgefragt wird. Trifft nicht zu: Es reicht nicht nur danach zu fragen ob ein Angehöriger dabei ist. Dem Patienten muss auch die Möglichkeit gegeben werden, den Angehörigen ins Zimmer dazu zu holen.
	S4 Augenkontakt herstellen (make eye contact) *generell kodieren* Trifft zu: Als Beobachter hat man den Eindruck, dass der Arzt keinen Blickkontakt vermeidet sondern nahezu durchgehenden aber nicht aufdringlichen Augenkontakt mit dem Patienten hat. Trifft nicht zu: Als Beobachter hat man den Eindruck, dass der Arzt: - Direkten Blickkontakt mit dem Patienten vermeidet - Nur flüchtigen Blickkontakt zulässt - Nur kurzzeitigen Blickkontakt hat, der aber nicht über längere Zeitabschnitte bestehen bleibt

	S5 Den Patienten begrüßen (greet the patient) Trifft zu: Als Beobachter hat man den Eindruck, dass der Arzt den Patienten freundlich begrüßt indem er z.B. folgendes tut: - Aufstehen um den Patienten zu begrüßen bzw. falls er schon sitzt, sich dem Patienten entgegen beugen - Händeschütteln und Augenkontakt - Sich mit Name und Position vorstellen. Trifft nicht zu: Als Beobachter hat man nicht den Eindruck, dass der Arzt eine angenehme Atmosphäre herstellt. Außerdem zeigt er <u>keine</u> der oben genannten Verhaltensweisen.
	S6 Den Patienten fragen wie es ihm geht um Interesse an dessen Wohlbefinden zu zeigen Trifft zu: Der Arzt fragt den Patienten wie er sich fühlt, z.B.: "Wie geht es Ihnen?" Kann auch kodiert werden wenn es später im Gespräch vorkommt. Trifft nicht zu: Der Arzt stellt keinerlei Fragen nach dem aktuellen Wohlbefinden des Patienten.
	S7 Den Patienten über eventuelle zeitliche Einschränkungen informieren (inform patient of time constraints there might be) Trifft zu: Der Arzt informiert den Patienten, dass er einen speziellen Zeitraum für ihn und die Besprechung reserviert hat und dass genug Zeit gegeben ist um die Diagnose zu diskutieren. Z.B.: "Ich habe jetzt eine halbe Stunde Zeit, so dass wir alles in Ruhe besprechen können." Trifft nicht zu: Der Arzt setzt keinerlei zeitlichen Rahmen für das Gespräch und / oder vermittelt ein Gefühl der Eile.
	S8 Ruhe sicherstellen (ensure silence: set pager/mobile phone on silent) Trifft zu: Der Arzt stellt sicher, dass die Besprechung nicht durch sein Mobiltelefon / seinen Piepser unterbrochen wird. Trifft nicht zu: Der Arzt hat sein Mobiltelefon und / oder seinen Piepser dabei und stellt ihn nicht auf lautlos.
P	**P1 Den Patienten fragen was er für den Grund der Untersuchung hält. (ask patient what he/she believes is the reason for which a consultation is made)** Trifft zu: bevor der Arzt die Diagnose übermittelt, bittet er den Patienten seine Symptome zu beschreiben um dadurch ein besseres Bild zu bekommen wie der Patient seine Lage wahrnimmt. Z.B.: „ Erzählen Sie mir bitte noch einmal wieso Sie denn eigentlich gekommen sind?" oder "In wie weit sind sie informiert über den Grund aus welchem die Kolloskopie bei Ihnen durchgeführt wurde?" Im Fall, dass dieser Gesprächaspekt bereits von dem Patienten selbst angesprochen worden ist nimmt der Arzt ihn auf und ist für den Inhalt

	empfänglich und reagiert darauf. Trifft nicht zu: Der Arzt erkundigt sich nicht nach oben genannten Aspekten. The doctor omits to ask such a question. Im Fall, dass dieser Gesprächaspekt bereits von dem Patienten selbst angesprochen worden ist, geht der Arzt nicht weiter darauf ein. “Hat man Ihnen schon irgendetwas gesagt?“ → reicht nicht aus!
	P2 Missverständnisse klären bzw. sicherstellen, dass die Patientenperspektive verstanden ist (clarify misunderstandings/ensure that one has understood the patient's perspective) Trifft zu: Dem Beobachter wird deutlich, dass der Arzt die Patientenperspektive verstanden hat - Wenn der Arzt die Informationen unter 'K' daran anpasst - Wenn der Arzt die Informationen, die er vom Patienten bekommen hat, paraphrasiert Trifft nicht zu: Dem Beobachter wird deutlich, dass der Arzt die Informationen nicht auf die Patientenwahrnehmung, speziell dessen Bedenken und Gedanken bzgl. der Diagnose anpasst.
	P3 Nach Gefühlen und Gedanken bezüglich des möglichen Ausgangs des Gesprächs fragen und herausfinden, ob der Patient Krankheitsverleugnung zeigt (ask about feelings and thoughts about the possible outcome of the consultation/check whether patient engages in illness denial) Trifft zu: Der Arzt setzt Fragen ein um die Patientenwahrnehmung der Situation, dessen Gefühle und Gedanken herauszufinden. Z.B.: „Das heißt sie haben eigentlich positive Erwartungen?“ oder „Haben sie sich da selber schon mal Gedanken gemacht von wo Ihre Beschwerden kommen könnten?“ Im Fall, dass dieser Gesprächaspekt bereits von dem Patienten selbst angesprochen worden ist nimmt der Arzt ihn auf und ist für den Inhalt empfänglich und reagiert darauf. “Wissen sie schon etwas darüber?” → trifft zu “Hat man Sie schon aufgeklärt was sein kann?” → trifft zu Trifft auch zu, wenn die Studenten erst später im Gespräch nachfragen, ob der Patient mit solch einer Diagnose gerechnet hatte oder was er vermutet hatte. Trifft nicht zu: Der Arzt stellt keinerlei Fragen um die Patientenperspektive der Situation, dessen Gefühle und Gedanken herauszufinden. Im Fall, dass diese Aspekte bereits von dem Patienten selbst angesprochen worden sind reagiert der Arzt nicht darauf und geht nicht weiter auf geäußerte Gefühle, Gedanken und Bedenken ein. “Hat man Ihnen schon etwas gesagt?“ → reicht nicht aus!
I	**I1 Herausfinden wie viele Informationen der Patient gerne hätte (inquire about the desired amount of information)** *Nur auf die Diagnose beziehen, nicht auf das Treatment!*

	Trifft zu: Bevor der Arzt die Diagnose verkündet benutzt er offene Fragen um herauszufinden ob der Patient die Testresultate eher sehr detailliert oder eher generell hören möchte. Z.B.: „Wie möchten Sie, dass ich Ihnen die Informationen gebe? Soll ich ins Detail gehen oder die Ergebnisse eher oberflächlich besprechen?“ Trifft nicht zu: Der Arzt beginnt sofort damit dem Patienten die Testresultate mitzuteilen ohne abzuwarten und zu checken wie viele Infos der Patient bereit ist aufzunehmen.
	I2 Anbieten alle zukünftig auftauchenden Fragen des Patienten zu beantworten (offer to answer any questions the patient might have in the future) Trifft zu: Der Arzt betont explizit, dass er bereit ist zukünftige Fragen des Patienten (am Telefon) zu beantworten. “Wenn Sie irgendwelche Fragen haben…” Trifft nicht zu: Der Arzt spricht den Aspekt zukünftiger Fragen nicht an. “Rufen Sie mich einfach an.”
K	**K1 Den Patienten warnen, dass die folgende Information wichtig und ernst ist (warn the patient that the information following is important and serious)** Trifft zu: Der Arzt beginnt mit der Besprechung der Testergebnisse indem er den Patienten warnt, dass er schlechte Nachrichten hat mit Aussagen wie z.B.: “Es ist leider so, dass ich Ihnen keine guten Nachrichten überbringen kann.” Trifft nicht zu: Der Arzt beginnt direkt damit die Diagnose zu nennen ohne eine vorbereitende Aussage die den Patienten darauf vorbereiten kann, dass er nun schlechte Nachrichten erhält.
	K2 Sprache und Vokabular auf das Verstehenslevel des Patienten anpassen (adjust language and vocabulary to patient’s level of comprehension) (am Ende noch kodieren) *“Anpassen“ ist nicht möglich, da die Studenten die Patienten nicht einschätzen können* Trifft zu: Der Arzt bereitet die Diagnose vor indem er erklärt, was während der Kolloskopie gemacht wurde, z.B.: dass eine Gewebeprobe entnommen und analysiert wurde und fährt dann fort zu erklären was gefunden worden ist. Das tut der Arzt indem er nicht-medizinisches Vokabular benutzt so wie „Gewebeprobe entnehmen” an Stelle von “Biopsie.” Der Arzt ist bereit Sätze umzuformulieren und andere Begriffe zu verwenden, falls der Patient etwas nicht verstanden hat. Es muss umgangssprachlich verständlich sein, sobald Fachbegriffe fallen müssen diese erklärt werden. Trifft nicht zu: Der Arzt benutzt viele medizinische Fachausdrücke und wird ungeduldig wenn der Patient nachfragt.
	K3 Unverblümte Offenheit vermeiden (avoid excessive bluntness) (am Ende noch kodieren)

	Nur in extremen Fällen nicht vergeben Trifft zu: Der Arzt verwendet keine dramatischen Statements wie “Wenn Sie sich nicht sofort einer Behandlung unterziehen werden Sie sterben.” Im Falle einer schlechten Diagnose benutzt der Arzt keine Aussagen wie “Es gibt nichts mehr was wir für Sie tun können.” Trifft nicht zu: Der Arzt beunruhig den Patienten indem er dramatische und alarmierende Statements benutzt wie “Sie haben sehr schlimmen Krebs und wenn Sie sich nicht sofort in Behandlung begeben werden Sie sterben.” oder “Es gibt nichts mehr was wir noch für Sie tun könnten.”
	K4 Information in kleinen Einheiten übermitteln (give small chunks of information) (am Ende noch kodieren) Trifft zu: Der Arzt überfordert den Patienten nicht mit zu viel Information sondern macht kruze Pausen um immer wieder mit Fragen herauszufinden ob der Patient alles verstanden hat. Der Arzt achtet auf Hinweise vom Patienten bzgl. des Wunschs nach mehr oder detaillierterer Informationen und fordert den Patienten dazu auf selbst Fragen zu stellen. Der Arzt gibt Informationen die sich auf die Bedenken des Patienten beziehen. Trifft nicht zu: Der Arzt hält einen Monolog ohne Pausen oder Nachfragen um das Verständnis beim Patienten zu sichern.
	K5 Diagnose geben (give diagnosis) Trifft zu: Der Arzt teilt dem Patienten die Diagnose mit indem er z.B. sagt “Sie haben Darmkrebs.” Trifft nicht zu: Der Arzt benennt die Krankheit nicht sondern benutzt Aussagen wie “Wir haben eine Auffälligkeit in Ihrem Darm gefunden und müssen jetzt schauen wie wir sie behandeln”
E	**E1 Emotionen des Patienten beobachten (observe patient’s emotions)** (am Ende noch kodieren) *Kodieren falls E2 oder E3 kodiert wurden* Trifft zu: Als Beobachter hat man den Eindruck, dass der Arzt den Gesichtsausdruck des Patienten und dessen Körpersprache aufmerksam studiert um jegliche emotionale Reaktionen bei ihm zu erkennen. Trifft nicht zu: Der Arzt achtet nicht auf emotionale Stresssignale in Gesichtsausdruck und Körpersprache des Patienten.
	E2 Emotionen identifizieren und benennen (identify and name the emotions) Trifft zu: Wenn der Arzt sich bezüglich der Art der Emotion die der Patient erlebt nicht sicher ist (z.B. wenn der Patient ganz ruhig ist) benutzt er offene Fragen um herauszufinden was der Patient denkt und fühlt. Z.B.: “Was geht denn jetzt in Ihrem Kopf vor?”

	„Wie fühlen Sie sich jetzt?“ „Was löst das jetzt für ein Gefühl in Ihnen aus?“ Trifft nicht zu: Der Arzt benennt werde die Emotionen des Patienten, noch versucht er sie herauszufinden oder erkennt sie an.
	E3 Grund für Emotion identifizieren (identify reason for the emotion) Trifft zu: Der Arzt fragt nach dem Grund warum der Patient z.B. geschockt oder verängstigt ist. “Haben Sie Angst weil Sie denken Sie brauchen einen künstlichen Darmausgang?” Trifft nicht zu: Es reicht nicht aus wenn der Patient von sich aus den Grund für seine Emotionen benennt. Der Arzt muss nachfragen.
	E4 Empathische und validierende Aussagen (empathic and validating statements) Trifft zu: Der Arzt last den Patienten wissen, dass er ihn und seine Gefühle versteht und mitfühlt. Z.B.: “Ich wünschte ich könnte Ihnen eine bessere Nachricht überbringen.” Der Arzt lässt den Patienten wissen, dass er versteht warum er aufgewühlt ist indem er eine Aussage macht, die sein Verständnis widerspiegelt. Z.B.: “Ich kann verstehen, dass Sie diese Nachricht schockiert”, “Das ist ganz natürlich, dass Sie dies so empfinden”. „Das klingt jetzt erstmal schlimm für Sie.“ Trifft nicht zu: Der Arzt geht nicht auf die emotionale Reaktion des Patienten ein.
	E5 Gesprächspausen / Ruhe zulassen und dem Patienten Zeit zum Antworten geben (allow for speaking pauses/silence/give patient time to respond) Trifft zu: Der Arzt lässt Momente der Stille zu und muss nicht die ganze Zeit sprechen. Dadurch gibt er dem Patienten Möglichkeit und Zeit seine Sätze zu beenden und nachzudenken. Pause muss auftreten, oder der Arzt unterbricht den Patienten nicht Trifft nicht zu: Der Arzt lässt keine Momente der Stille zu und lässt den Patienten seine Sätze nicht beenden bzw. ergreift das Wort selbst wieder zu schnell.
Z	**Z1 Herausfinden, ob der Patient bereit und in der Lage ist den Behandlungsplan zu besprechen (check patients ability/readiness to discuss treatment plan)** Trifft zu: Der Arzt versucht herauszufinden ob der Patient in diesem Moment bereit ist den Behandlungsplan zu besprechen oder ob er für jetzt erst einmal genug gehört hat. Z.B.: “Wollen Sie die Informationen in Ruhe verarbeiten oder sollen wir zusammen schon einmal schauen wie es dann weiter geht?“ Der Arzt bietet an einen weiteren Besprechungstermin zu vereinbaren. Trifft nicht zu: Der Arzt beginnt direkt damit die einzelnen Schritte des Behandlungsplans darzulegen.
	Z2 Den Patienten in den Entscheidungsprozess einbeziehen (involve patient into

the decision making process)

Trifft zu: Der Arzt macht dem Patienten deutlich, dass er bei den nächsten Schritten die Wahl hat. Z.B.: "Was denken Sie denn wie wäre es am besten für sie fortzufahren?"
„Ich kann Ihnen anbieten, dass wir darüber reden und ich Ihnen die Möglichkeiten aufzeige."
„wir"-Form ist ausreichend aber nur wenn sich das "wir" auf den Patienten und nicht das Behandlerteam alleine bezieht.

Trifft nicht zu: Der Arzt trifft die Entscheidung für die anstehende Behandlung ohne die Meinung des Patienten einzuholen.

Z3 Überprüfen inwieweit der Patient den Behandlungsplan verstanden hat und wie realistisch seine Erwartungen für das Endergebnis sind (check for understanding of the treatment plan and patient's accurate/realistic expectations of the outcome)

Trifft zu: Der Arzt achtet auf Aussagen des Patienten, die auf eine falsche Wahrnehmung des Behandlungsplans und unrealistische Erwartungen bezüglich des Endergebnisses hinweisen können. Der Arzt ist bereit sich noch einmal die Bedenken und Ängste des Patienten anzuhören.

Trifft nicht zu: Der Arzt zeigt keine der oben beschriebenen Verhaltensweisen.

Z4 Missverständnisse korrigieren (correct any misunderstandings)

Trifft zu: Der Arzt korrigiert Fehlannahmen, Ängste aber auch unrealistische Hoffnungen die vom Patienten geäußert werden indem er z.B. relaitiviert: „Ja, sie können theoretisch an Krebs sterben, aber in Ihren Fall ist das sehr unwahrscheinlich".
Missverständnisse können sein: "Muss ich sofort operiert werden?", „Muss ich jetzt direkt ins Krankenhaus?"

Trifft nicht zu: Der Arzt korrigiert die vom Patienten geäußerten falschen Annahmen, Ängste und unrealistischen Hoffnungen nicht.

Z5 empathisch reagieren (react empathically) *(wird immer automatisch kodiert sobald E4 kodiert wird)*

Trifft zu: Der Arzt gibt dem Patienten die Möglichkeit seine Sorgen noch einmal zu äußern und reagiert mit empathischen Aussagen wenn der Patient sich aufregt (see STEP 5). Hoffnung sollte nur soweit gegeben werden wie sie auch tatsächlich realistisch erreichbar ist.

Trifft nicht zu: Der Arzt geht in keiner Weise auf die Emotionen des Patienten ein. Ein Nicken, während der Patient etwas erzählt, reicht nicht aus.

Z6 Persönliche Ziele des Patienten verstehen (understand patient's specific goals)

Trifft zu: Der Arzt achtet auf Ziele des Patienten (z.B. eine lange geplante Reise oder die familiäre Situation).
Dieser Aspekt trifft zu sobald etwas ein persönliches Anliegen des Patienten ist. Also auch bei z.B. Besuch eines Heilers, Kinder zu Hause, Plan eine Reise zu machen, Vermeidung eines künstlichen Darmausgangs weil dieser nicht zur Lebensführung passt

Trifft nicht zu: Der Arzt ignoriert die Ziele des Patienten.

Z7 Ein Taxi rufen bzw. die Familie kontaktieren (call a cab/family)

Trifft zu: wenn der Arzt zumindest fragt, wie der Patient nach Hause kommt. Zum Beispiel bietet der Arzt an ein Taxi zu rufen oder ein Familienmitglied zu kontaktieren damit der Patient nicht selbst mit dem Auto nach Hause fahren muss. Z.B.: "Soll ich Ihnen ein Taxi rufen?" oder „Vielleicht rufen wir ihren Mann an, dass er Sie abholen kommen soll."

Trifft nicht zu: Der Arzt bietet nicht an ein Taxi zu rufen und ist auch nicht daran interessiert wie der Patient sicher nach Hause kommt. Er lässt den Patienten selbst nach Hause fahren.

Z8 Behandlungsplan besprechen (discuss treatment plan)

Trifft zu: Der Arzt bespricht den Behandlungsplan mit dem Patienten: "Wir vereinbaren jetzt einen nächsten Termin, dort müssen wir weitere Untersuchungen durchführen."

Trifft nicht zu: Der Arzt bespricht den Behandlungsplan nicht mit dem Patienten.

Appendix 7: Items of the scale to measure affect related to breaking bad news.

Bitte geben Sie an, inwiefern Sie den folgenden Aussagen zum Überbringen von schlechten Nachrichten zustimmen.

	Stimme nicht zu 1	2	teils teils 3	4	Stimme zu 5
Das Überbringen von schlechten Nachrichten ist aus meiner Sicht völlig unproblematisch.	○	○	○	○	○
Es ist mir unangenehm eine schlechte Nachricht zu überbringen.	○	○	○	○	○
Ich habe eine hohe Hemmschwelle eine schlechte Nachricht zu überbringen.	○	○	○	○	○
Ich habe Angst davor eine schlechte Nachricht zu überbringen.	○	○	○	○	○

Absenden

Appendix 8: Difficulty rating of reading the spikes text and working with the video-based modeling worked example.

Bitte beantworten Sie kurz die folgenden Fragen

Bitte geben Sie an wie schwer Text zu den SPIKES gewesen ist, indem Sie die entsprechende Antwort ankreuzen.
Die Skala reicht von 1(sehr einfach) bis 9(sehr schwer).

sehr einfach 1	ziemlich einfach 2	einfach 3	eher einfach 4	mittelschwer 5	eher schwer 6	schwer 7	ziemlich schwer 8	sehr schwer 9
○	○	○	○	○	○	○	○	○

Bitte geben Sie an wie schwer diese Lerneinheit (das letzte lange Video/Phase 2) gewesen ist, indem Sie die entsprechende Antwort ankreuzen.
Die Skala reicht von 1(sehr einfach) bis 9(sehr schwer).

sehr einfach 1	ziemlich einfach 2	einfach 3	eher einfach 4	mittelschwer 5	eher schwer 6	schwer 7	ziemlich schwer 8	sehr schwer 9
○	○	○	○	○	○	○	○	○

Appendix 9: Overview of Experts' comments included in the three video-based worked examples.

	Video-case A Coping Student Model	Video-case B Mastery Student Model	Video-case C Mastery Professor Model
Gap with 1st self-explanation scaffold	00:00 bis 00:18	00:45	00:38
Experts‘ comment after 1st self-explanation scaffold	Sequenz 2 **Gut:** Der Arzt fragt den Patienten nach seinen Beschwerden, die zur Untersuchung geführt haben. Er bestärkt ihn durch Nicken und fragt genauer nach. Das ist gut, da der Arzt nur auf diesem Weg die Sicht des Patienten kennenlernen kann. Durch das Nicken fühlt sich der Patient ernst genommen und bestärkt seinen eigenen Eindruck zu schildern.	Sequenz 2 **Gut:** Der Arzt fragt die Patientin nach ihren Beschwerden, die zur Untersuchung geführt haben. Er bestärkt sie durch Nicken, fasst die Informationen der Patientin zusammen und fragt genauer nach. Das ist gut, da der Arzt nur auf diesem Weg die Sicht der Patientin kennenlernen kann. Durch das Nicken fühlt sich die Patient ernst genommen und bestärkt ihren eigenen Eindruck zu schildern.	Sequenz 2 Gut: Der Arzt fragt die Patientin nach ihren Beschwerden, die zur Untersuchung geführt haben. Er bestärkt sie durch Nicken, fasst die Informationen der Patientin zusammen und fragt genauer nach. Das ist gut, da der Arzt nur auf diesem Weg die Sicht der Patientin kennenlernen kann. Durch das Nicken fühlt sich die Patient ernst genommen und bestärkt ihren eigenen Eindruck zu schildern. Enthält: Spikes P1-P2
Gap with 2nd self-explanation scaffold	02:51 bis 03:30	03:24 bis 03:36	01:16
Experts‘ comment after 2nd self-explanation scaffold	Sequenz Der Patient reagiert heftig auf die Aussage, dass er operiert werden soll. Der Arzt geht aber nur ansatzweise auf die Nachfrage "Ach, ich soll operiert werden?!" ein. Stattdessen gibt der Arzt weitere Informationen und geht	Sequenz 67 Gut: Der Arzt bekommt von der Patientin signalisiert, dass sie für die Beschäftigung mit dem Behandlungsplan bereit ist. So kann er das Gespräch der Patientin entsprechend dem Bedürfnis zur	Sequenz 8 Der Arzt vergewissert sich, dass keine Missverständnisse auf Seiten der Patientin über die Erkrankung und den Behandlungsplan vorliegen. Z3 (siehe 07:14 OT, Sequenz 9 TH)

	wieder auf den Behandlungsplan ein. **Fehlt:** Die Emotionen (Angst, Furcht) des Patienten werden durch den Arzt nicht identifiziert und benannt. Dadurch findet keine Identifizierung der Gründe für die Emotionen statt. Ein verbindendes Statement wie „Ich kann Ihre Angst gut nachvollziehen“ wäre angemessen gewesen. Dadurch hätte der Arzt dem Patienten in dessen Schock Unterstützung und Solidarität demonstriert.	Gestaltung der nächsten Zeit anpassen. Z1 (Sequenz 67 Siebeck) Gut: An dieser Stelle geht der Arzt zunächst auf die Emotionen der Patientin ein. Er macht ein verbindendes Statement und signalisiert damit Solidarität und Unterstützung. Z5 (siehe Sequenz 5 Siebeck mit „Körpersprache & Wiederholen“) Das Statement „ich möchte Sie nicht mit Therapievorschlägen überfahren“ oder „ich kann Sie verstehen wenn Sie erst einmal Zeit für sich brauchen“ mit anschließender Pause ist in dieser Situation sehr passend. Z2 (siehe Sequenz 10) Der Arzt erfasst die Aussagen der Patientin und gibt ihr Zeit und Ruhe um sich an die Diagnose zu gewöhnen um danach das Gespräch fortsetzen zu können. Hierbei ist es hilfreich wenn der Arzt auch einmal schweigt. Er überschüttet sie nicht mit Informationen zum weiteren Vorgehen sondern identifiziert die zugrunde liegenden Emotionen. (Siebeck Sequenz 4)	
Gap with 3rd self-	04:38 bis 07:14	04:01 bis 04:03	05:10

explanation scaffold	Sequenz	Sequenz 5	Sequenz 4
Experts' comment after 3rd self-explanation scaffold	Der Patient ist an dieser Stelle zum ersten Mal mit der Diagnose „Darmkrebs" konfrontiert und reagiert emotional. **Fehlt:** An dieser Stelle müsste der Arzt zunächst auf die Emotionen des Patienten eingehen, ein verbindendes Statement machen und damit Solidarität und Unterstützung signalisieren. Das Statement „Ich kann Sie verstehen, dass Sie jetzt erstmal überfahren sind" mit anschließender Pause ist sehr gut, kommt aber zu spät. Er missversteht die Aussagen des Patienten, identifiziert nicht die zugrunde liegenden Emotionen und gibt lediglich weitere Informationen um ihn zu beruhigen. Diese sind an dieser Stelle teilweise viel zu detailliert als dass der Patient sie aufnehmen könnte („Der Krebs ist 30cm von After entfernt") und bewirken eher eine zusätzliche Überforderung als dass sie dem Patienten in seiner aktuellen Lage helfen würden. **Besser** wäre es, dem Patienten Zeit und Ruhe zu geben sich an die Diagnose zu gewöhnen um danach das	Gut: Der Arzt macht eine klare Aussage indem er die schwere Diagnose sagt. Er lässt die Patientin nicht länger als unnötig über die Ergebnisse der Untersuchung im Unklaren. K5 Die Patientin wird an dieser Stelle zum ersten Mal mit der Diagnose „Darmkrebs" konfrontiert und reagiert emotional. Sie ist über die Diagnose sehr überrascht da sie zur Hämorrhoiden Abklärung gekommen ist und nicht mit einer Krebsdiagnose gerechnet hatte.(siehe 04:38 OT, siehe Sequenz 4 Siebeck) Mittelmäßig: Die Emotionen der Patientin (Angst, Furcht) werden durch den Arzt anhand von empathisch bestätigenden Aussagen weitestgehend identifiziert und benannt. Der Arzt könnte genauer auf die Emotionen eingehen.E2 Gut: Das verbindende Statement „auch ich hätte mir gewünscht, dass ich Ihnen eine andere Diagnose hätte mitteilen können" ist angemessen. Hierdurch kann der Arzt der Patientin in ihrem	Die Patientin wird an dieser Stelle zum ersten Mal mit der Diagnose „Darmkrebs" konfrontiert und reagiert emotional. E3 Sie ist über die Diagnose sehr überrascht da sie nicht mit einer Krebsdiagnose gerechnet hatte. P3 (siehe 04:38 OT, Sequenz 5 TH) Der Arzt erfasst die Aussagen der Patientin und gibt ihr Zeit und Ruhe um sich an die Diagnose zu gewöhnen um danach das Gespräch fortsetzen zu können. Z4 Hierbei ist es hilfreich wenn der Arzt auch einmal schweigt. E5 Er überschüttet sie nicht mit Informationen zum weiteren Vorgehen Z3 sondern identifiziert die zugrunde liegenden Emotionen und benennt sie. E2 (Sequenz 67, 4 TH) Gut: Das verbindende Statement „da müssen sie schlucken" und „das setzt Ihnen richtig zu" ist angemessen. Hierdurch kann der Arzt der Patientin in ihrem Schock Unterstützung und Solidarität demonstrieren. E4 (siehe Sequenz 5 TH) Stattdessen wird angezeigt:

	Gespräch fortsetzen zu können. Dabei kann es hilfreich sein, wenn der Arzt auch einmal schweigt.	Schock Unterstützung und Solidarität demonstrieren.(siehe Sequenz 4 Siebeck)	Gut: Der Arzt fragt die Patientin nach ihren Beschwerden, die zur Untersuchung geführt haben. Er bestärkt sie durch Nicken, fasst die Informationen der Patientin zusammen und fragt genauer nach. Das ist gut, da der Arzt nur auf diesem Weg die Sicht der Patientin kennenlernen kann. Durch das Nicken fühlt sich die Patient ernst genommen und bestärkt ihren eigenen Eindruck zu schildern.
Gap with 4th self-explanation scaffold	08:16:00	06:59	06:38
Experts' comment after 4th self-explanation scaffold	**Gut:** Der Arzt gibt eine klare Antwort und relativiert anschließend. Anschließend verliert er sich allerdings in **Behandlungsoptionen** anstatt bei der Frage des Patienten zu bleiben. Diese zusätzlichen Informationen sind nicht auf die Bedürfnisse des Patienten angepasst, können ihn verwirren und sind an dieser Stelle des Gesprächs deswegen nicht hilfreich. (siehe 06:31 rechts) Der Patient sorgt sich und drückt dies mit seiner **Frage**, ob er sterben wird, aus. **Fehlt:** Der Arzt sollte auf diese Sorge	Sequenz 11a Gut: Der Arzt geht auf die Situation und Bedürfnisse der Patientin ein indem er ihr den **Behandlung**splan und weitere notwendige Untersuchungen erklärt.(siehe Siebeck Sequenz 8) **Gut:** Der Arzt benutzt angemessenes Vokabular und keine Fachausdrücke. Insgesamt ist die Aussage auf das Verständnis der Patientin angepasst. Dadurch wird gewährleistet, dass die Patientin eine Chance hat alles zu verstehen. Nur so können Arzt und Patientin sich gleichberechtigt	Sequenz 10 Der Arzt geht mit Statements wie „wir wollen den Tumor nicht munter weiter wachsen lassen“ und „dann machen wir den Termin aus“ auf die Patientin ein und bindet sie in den weiteren Entscheidungsprozess ein. Z2 (siehe Sequenz 67 TH) Gut: An die Notwendigkeit eines nächsten Termins wird gedacht. (siehe 10:44 OT)

	eingehen und mit dem Patienten über seine Emotionen sprechen und nicht ausschließlich über Behandlungsmöglichkeiten. **Schlecht:** Die Aussage, dass das Ausgehen der Haare nicht so ein großes Problem sei, ist fehl am Platz. Der Arzt sollte die Sorgen und Bedenken seines Patienten ernst nehmen und auf ihn eingehen.	miteinander über die Diagnose und das weitere Vorgehen unterhalten. K2 (Siebeck Sequenz 8)	

Appendix 10: Items of the anxiety scale applied in the 2nd study.

angst01	Vor dem Überbringen schlechter Nachrichten habe ich Angst.[35]
angst02	Bevor ich die schlechte Nachricht überbringe, bin ich beunruhigt und fühle mich unwohl.
angst03	Bevor ich die schlechte Nachricht überbringe, bin ich sehr nervös.
angst04	Ich gerate während des Überbringens schlechter Nachrichten schnell in Panik.
angst05	Ich mache mir Sorgen, ob ich das schwierige Gespräch führen kann.
angst06	Vor lauter Nervosität würde ich am liebsten das Überbringen der schlechten Nachricht an einen Kollegen delegieren.
angst07	Ich bin so aufgeregt, dass ich denke: Wenn es bloß vorbei wäre.
angst08	Ich habe so viel Angst, dass ich mir wünsche, weit weg zu sein.
angst09	Wenn ich an das schwierige Gespräch denke, bekomme ich ein flaues Gefühl im Magen.
angst10	Wenn das schwierige Gespräch beginnt, habe ich Herzklopfen.
angst11	Bevor ich die schlechte Nachricht überbringe, habe ich ganz zittrige Hände.

Appendix 11: Inhibition scale used in the 2nd study in comparison to scale used in the 1st study.

	Change
Das Überbringen schlechter Nachrichten ist aus meiner Sicht völlig unproblematisch	
Es ist mir unangenehm, ***Patienten*** schlechte Nachrichten zu überbringen	Explicit hint to patients added
Ich habe eine hohe Hemmschwelle schlechte Nachrichten***, z.B. eine Krebsdiagnose,*** zu überbringen	
Das Überbringen einer schlechten Diagnose bereitet mir keine Probleme	New items
Ich versuche, schwierige Gespräche mit Patienten möglichst kurz zu halten	
Ich schiebe schwierige Gespräche mit Patienten möglichst lange heraus	
Ich versuche, schwierige Gespräche mit Patienten schnell hinter mich zu bringen	

[35] Formulation of the item in the 1st study: ‚Ich habe Angst davor eine schlechte Nachricht zu überbringen'.

Appendix 12: E-Skala for self-assessment of empathy by Leibetseder and colleagues (2001).

Sie werden jetzt eine Reihe von Aussagen lesen, die jeweils bestimmte menschliche Eigenschaften oder Reaktionen beschreiben, die etwas mit Gefühlen zu tun haben.
Bitte kennzeichnen Sie auf der Antwortskala, inwieweit die Aussage auf Sie zutrifft.

	Trifft nicht zu	Trifft eher nicht zu	Weder noch	Trifft eher zu	Trifft zu
Ich finde es etwas übertrieben sich in Bücher oder Filme hineinzusteigern	☐	☐	☐	☐	☐
In die Gefühle von Romanfiguren lebe ich mich richtig hinein	☐	☐	☐	☐	☐
Wenn ich einen Film sehe, stelle ich mir oft vor, wie es mir selbst anstelle der betreffenden Person erginge	☐	☐	☐	☐	☐
Wenn ich eine interessante Geschichte lese, stelle ich mir vor, wie ich wohl in so einer Situation zurecht käme	☐	☐	☐	☐	☐
Bei einem guten Film kann ich mich sehr leicht in den Hauptsdarsteller hineinversetzen	☐	☐	☐	☐	☐
Bei einer interessanten Erzählung stelle ich mir vor, wie es mir dabei erginge	☐	☐	☐	☐	☐
Wenn jemand in einem Fernsehquiz Geld gewinnt, stelle ich mir sofort vor, wie ich mich an seiner/ihrer Stelle fühlen würde	☐	☐	☐	☐	☐
Nach einem Theaterstück oder nach einem Film fühle ich mich teilweise so, als ob ich selbst einer der Charakteren wäre	☐	☐	☐	☐	☐
Es passiert mir eher selten in einem guten Buch oder in einem guten Film besonders aufzugehen	☐	☐	☐	☐	☐
Ich kann sehr leicht die Gefühle von Romanfiguren nachempfinden	☐	☐	☐	☐	☐
Wenn ich einen interessanten Roman lese, stelle ich mir vor, wie ich mich fühlen würde, wenn mir diese Ereignisse passieren würde	☐	☐	☐	☐	☐
Ich neige dazu, Theaterstücke oder Filme derart mitzuerleben, dass ich empfinde, als wäre ich selbst eine der handelnden Personen	☐	☐	☐	☐	☐
Wenn ich einen guten Film ansehe, kann ich sehr leicht die Hauptdarsteller nacherleben	☐	☐	☐	☐	☐
Der Anblick weinender Menschen bringt mich aus der Fassung	☐	☐	☐	☐	☐

Ich neige dazu, mich in die Probleme eines Freundes hineinzuleben	☐	☐	☐	☐	☐
Wenn ich einen sehr alten Menschen sehe, frage ich mich oft, wie ich mich an seiner/ihrer Stelle fühlen würde	☐	☐	☐	☐	☐
Filme über Krieg und töten regen mich innerlich auf	☐	☐	☐	☐	☐
Es beunruhigt mich mehr als die meisten anderen Menschen, wenn ich sehe, wie sich ein Freund verletzt	☐	☐	☐	☐	☐
Ich bin oft ziemlich berührt durch Dinge, die vor meinen Augen passieren	☐	☐	☐	☐	☐
Ich fühle oft Betroffenheit und Mitgefühl mit anderen Menschen, die weniger glücklich sind als ich	☐	☐	☐	☐	☐
Missgeschicke anderer Menschen berühren mich meist nicht sehr	☐	☐	☐	☐	☐
Manchmal versuche ich meine Freunde dadurch besser zu verstehen, indem ich mir die Dinge aus ihrer Sicht vorstelle	☐	☐	☐	☐	☐
Wenn ich ein behindertes Kind sehe, versuche ich mir vorzustellen, wie es sich in bestimmten Situationen fühlt	☐	☐	☐	☐	☐
Es macht mich traurig, in einer Gruppe einen einsamen Menschen zu sehen	☐	☐	☐	☐	☐
Die Menschen um mich haben einen großen Einfluss auf meine Stimmung	☐	☐	☐	☐	☐

Appendix 13: PSI scale based on Process Scales by Hartmann & Schramm (2008).

Cognitive PSI	psi01	Ich habe genau beobachtet, wie sich der Arzt verhalten hat.
	psi02	Ich habe versucht, mir einen Reim auf die Handlungen des Arztes zu machen.
	psi03	Ich habe mich hin und wieder gefragt, ob ich schon einmal in ähnlichen Situationen wie der Arzt war.
	psi04	Ich habe immer wieder eingeschätzt, ob ich die Äußerungen oder das Verhalten von dem Arzt schlecht oder gut finde.
	psi05	Ich habe eigentlich nie darüber nachgedacht, was der Arzt als Nächstes tun oder sagen könnte.
	psi06	Ich habe bei der Beobachtung von dem Arzt immer wieder überlegt, ob ich selbst die Sache besser oder schlechter gemacht hätte als er.
Affective PSI	psi07	Ich war immer voller Mitgefühl mit dem Arzt.
	psi08	Ich hätte stets sagen können, wie sich der Arzt in verschiedenen Situationen gefühlt hat.
	psi09	In einigen Situationen war mir, als ob ich die gleichen Gefühle durchlebe wie der Arzt.
	psi10	Ich hatte den Eindruck, dass sich die Stimmung von dem Arzt kaum auf mich übertragen hat.
	psi11	Die Gefühle, die ich bei dem Arzt beobachtete, haben sich nicht auf mich übertragen.
	psi12	Was der Arzt gesagt oder getan hat, löste keinerlei Emotionen bei mir aus.
	psi13	Ich reagierte eher nüchtern und emotional unbeeindruckt auf den Arzt.

Appendix 14: Items to measure cognitive load based on the scale by Opfermann (2008).

Intrinsic Load	Wie leicht oder schwer finden Sie (im Moment) das Überbringen schlechter Nachrichten?
Extraneous Load	Wie leicht oder schwer fällt es Ihnen, mit der Lernumgebung (den Videos, den Expertenlösungen, etc.) zu arbeiten?
	Wie leicht oder schwer fällt es Ihnen, in der Lernumgebung wichtige und unwichtige Informationen zu unterscheiden?
	Wie leicht oder schwer fällt es Ihnen, alle Informationen, die Sie benötigen, in der Lernumgebung zu finden?