Reflexivity in Teams

Dissertation zur Erlangung des Doktorgrades der Philosophischen Fakultät der Christian-Albrechts-Universität zu Kiel

vorgelegt von

Kai-Philip Otte

Kiel

28.02.2017

Erstgutachter: Prof. Dr. Udo Konradt

Zweitgutachter: Prof. Dr. Jens Möller

Tag der mündlichen Prüfung: 14.06.2017

Durch den zweiten Prodekan, Prof. Dr. Elmar Eggert,

zum Druck genehmigt am: 14.06.2017

„Wenn nun der wahrnimmt, der sieht, dass er sieht, und hört, dass er hört, und als Gehender wahrnimmt, dass er geht, und wenn es bei allem anderen ebenso eine Wahrnehmung davon gibt, dass wir tätig sind, so dass wir also wahrnehmen, dass wir wahrnehmen, und denken, dass wir denken: und dass wir wahrnehmen und denken, ist uns ein Zeichen, dass wir sind (…)."

Aristoteles

Zusammenfassung

Kern dieser Dissertation ist das psychologische Konzept *Team Reflexivität*. Team Reflexivität kann definiert werden als das Ausmaß, innerhalb dessen Teams kollektiv über Strategien und Prozesse reflektieren, und sie gemäß den gewonnenen Erkenntnissen anpassen. Die zentrale Annahme des Konstruktes ist dabei, dass Team Reflexivität die Leistung eines Teams steigert. Entgegen der augenscheinlichen Plausibilität des Konzepts, hat die Reflexivitätsforschung der letzten zwei Jahrzehnte allerdings überraschenderweise ein relativ heterogenes Bild hinsichtlich des prognostizierten Zusammenhangs hervorgebracht, darunter neutrale und sogar negative Beziehungen zwischen Team Reflexion auf Leistung. Das übergeordnete Ziel dieser Dissertation ist daher die Identifizierung von möglichen Gründen und Ursachen für diese heterogenen Ergebnisse.

Die erste Studie der Dissertation setzt sich mit dem Problem der Konzeptualisierung und der Messung von Team Reflexivität auseinander. Zuerst wird daher der Begriff Team Reflexivität nicht weiter wie zuvor als Bezeichnung für ein einzelnes Omnibuskonstrukt verwendet, sondern für ein Framework bestehenden aus mehreren, reflexiven Prozessen. Unter ihnen befindet sich der Team Reflexionsprozess, der im Fokus dieser Arbeit stehen wird. Es wird argumentiert, dass ein wichtiger Grund für die berichtete, heterogene Befundlage, in den bisher verwendeten Fragebögen zur Messung von Team Reflexion liegt. Insbesondere liegen hier Probleme hinsichtlich der Inhalts- und Konstruktvalidität vor. Darüber hinaus wurden mit bisherigen Messmethoden lediglich quantitative Elemente betrachtet, das heißt, wie häufig bestimmte Verhaltensweisen aufgetreten sind. Relevante qualitative Elemente sind dadurch nicht erfasst worden, beispielsweise die Tiefe und Detailliertheit von Team Reflexion. Aus diesem Grund wird in der ersten Studie ein neuer Fragebogen entwickelt und validiert, der neben der Quantität auch die Qualität von Reflexion abbildet. In vier ausführlichen Validierungsuntersuchungen mit insgesamt 803 Studierenden

und Arbeitnehmern kann die Validität und Effektivität der neuen Skala in verschiedenen Kontexten, sowie hinsichtlich diverser Kriterien (z.B. Leistung) und Verhaltensweisen dargelegt werden.

Die zweite Studie beschäftigt sich auf Basis der neu entwickelten Skala mit tiefergehenden Fragen zur Beziehung zwischen Team Reflexion und Leistung. Speziell wird dabei auf die Kritik eingegangen, dass sich die bisherige Reflexivitätsforschung vergleichsweise selten mit längsschnittlichen Zusammenhängen und Entwicklungen auseinandergesetzt, und damit den wichtigen Aspekt der Zeit außeracht gelassen hat.

Aus diesem Grund wird mittels einer umfangreichen Laboruntersuchung mit 47 Teams und zwei Messzeitpunkten untersucht, inwiefern sich Feedback zu vorangehender Leistung gemäß selbst- und team-regulatorischen Annahmen auf die Reflexion auswirkt, und wie quantitative und qualitative Reflexion mit Leistungsverbesserungen von Teams in Beziehung stehen. Zusätzlich zur fragebogenbasierten Methode wird ein neues, verhaltensbasiertes Maß zur Messung von Team Reflexion überprüft und mit dem Fragebogenverfahren verglichen. Darüber hinaus wird schließlich der Implementierungsprozess betrachtet, das heißt die Umsetzung von Strategien, die auf Basis der aus der Reflexion gewonnenen Erkenntnissen entwickelt wurden, und somit die Effekte von Reflexion auf Leistung vermitteln soll. Die Ergebnisse zeigen, dass vorangehende Leistung und Team Reflexion gemäß der selbst- und team-regulatorischen Annahmen negativ zusammenhängen. Darüber hinaus wird gezeigt, dass quantitative Team Reflexion im Rahmen einer zeitlich begrenzten Reflexionsphase negativ mit Leistungsverbesserung zusammenhängt, während qualitative Reflexion eine positive Beziehung aufweist. Interessanterweise konnten keine Hinweise hinsichtlich der prognostizierten, vermittelnden Rolle von Implementation gefunden werden. Insgesamt liefert die zweite Studie daher erstmals relevante und neue Einsichten hinsichtlich der mehrdeutigen

Beziehung zwischen Reflexion und Leistung, und deutet über die in Studie 1 gewonnenen Erkenntnisse hinaus auf die Validität des entwickelten Fragebogens hin.

In der dritten Studie wird die Frage nach der zeitlichen Komponente als wichtiger Einflussfaktor auf den Zusammenhang zwischen Reflexion und Leistung tiefergehender untersucht. Aus diesem Grund wird Team Reflexion im Kontext der Punctuated Equilibrium Theorie (Theorie des punktierten Gleichgewichts) betrachtet. Kernelement der Theorie ist die evolutionstheoretisch gefußte Annahme, dass sich organisationale Entwicklungs- und Veränderungsprozesse im Rahmen von langen, stabilen Perioden abspielen (Equilibrium), die von kurzen und revolutionären Perioden unterbrochen werden, innerhalb derer Teams gezwungen sind sich fundamental anzupassen. Die Theorie besitzt zwar organisationstheoretisch einen großen Stellenwert, wurde in der organisationspsychologischen Praxis allerdings nur in einem sehr begrenzten Umfang validiert und eingesetzt. Als mögliche Ursache hierfür wird in dieser Studie argumentiert, dass die Theorie aufgrund ihres makroskopischen Fokus‘ auf die strukturelle Komponente der Veränderung in Form der beiden Periodentypen (Equilibrium und Revolution) wichtige prozesstheoretische Aspekte nicht berücksichtigt hat, weswegen die Anwendbarkeit der Theorie auf der Teamebene, trotz ihrer Relevanz, begrenzt war. Darüber hinaus wird argumentiert, dass sich diese Relevanz für die Betrachtung von Teamprozessen vor allem daraus ergibt, dass die Effekte der Prozesse auf relevante Ergebniskriterien (z.B. Leistung) von den Anforderungen abhängen, die aus den Periodentypen der Punctuated Equilibrium Theorie resultieren. Daher wird in dieser Studie untersucht, inwiefern diese Abhängigkeit von zeitlich variierenden Kontextfaktoren auch zur Mehrdeutigkeit der Zusammenhänge zwischen Team Reflexion und Leistung beiträgt.

Zuerst wird die Theorie daher um mehrere prozesstheoretische Elemente erweitert und anschließend im Rahmen eines komplexen Wirtschaftsplanspiels mit 86 Teams über 6 Messzeitpunkte angewandt. Ein wichtiges Element des Planspiels ist ein wirtschaftlicher

Einbruch nach dem dritten Messzeitpunkt, der die Betrachtung von Revolutions- und anschließenden Equilibriumsperioden erlaubt. Die Ergebnisse zeigen, dass die getroffenen Annahmen hinsichtlich der Kontextabhängigkeit relevant sind, und sich der Zusammenhang zwischen Leistung und den verschiedenen, reflexiven Prozessen, die mit der in Studie 1 entwickelten Skala erfasst werden, systematisch über die Zeit ändert. Darüber hinaus stellt diese Studie auch eine der ersten, organisationspsychologischen Anwendungen der Punctuated Equilibrium Theorie dar.

Zum Abschluss der Dissertation werden schließlich die Ergebnisse der drei Studien diskutiert, und Implikationen für Theorie und Praxis erläutert.

Inhaltsverzeichnis

Kapitel 1

Einleitung und theoretischer Rahmen

1.1 Einleitung

Der Begriff *Philosophie* lässt sich aus dem griechischem mit ‚Liebe zum Wissen' übersetzen, wird aber auch oft als das ‚Streben nach Erkenntnis' bezeichnet (Röd, 2000). Ein elementarer Eckpfeiler dieses Strebens nach Wissen und Erkenntnis ist die Reflexion, nach Platon auch die ‚Erkenntnis der Erkenntnis', und nach Aristoteles das ‚Denken des Denkens'. Etwas nüchterner fasst der Duden (2017) den Begriff Reflexion als ‚Nachdenken, Überlegung, prüfende Betrachtung' auf.

Entgegen der schon weit über zweitausend Jahre existierenden Reflexionstradition in der Philosophie, innerhalb der eine Wandlung vom griechischen *epostrophé* hin zum heutigen Begriff der *Reflexion* stattfand, ist die systematische Beschäftigung mit der Reflexion in der Arbeits- und Organisationspsychologie vergleichsweise jung. Häufig zitiert wird dabei Schön (1983), der in seinem Buch ‚The Reflective Practitioner' zur Reflexion festhielt:

> Both ordinary people and professional practitioners often think about what they are doing, sometimes even while doing it. Stimulated by surprise, they turn thought back on action and on the knowing which is implicit in action (Schön, 1983, S. 50).

In der Arbeits- und Organisationspsychologie entwickelten sich schließlich mehrere Reflexions-Forschungszweige, unteranderem vergleichsweise einfache ‚Stop and think' Interventionen (Okhuysen, 2001), über vor allem im Militär verbreitete und stark strukturierte ‚Debriefing' oder ‚After-Action-Review' Techniken (Smith-Jentsch, Cannon-Bowers, Tannenbaum, & Salas, 2008; Tannenbaum & Cerasoli, 2013), hin zu komplexeren Team Reflexionsmodellen im organisationalen Alltag (West, 2000). Trotz dieser Vielfalt teilen alle Ansätze allerdings die Grundannahme, dass sich Reflexion positiv auf Kriterien, wie zum Beispiel die Leistung auswirkt. In dieser Dissertation wird dabei insbesondere auf den

Zusammenhang zwischen Reflexion und Leistung in *Teams* fokussiert. Teams sind aufgrund der Komplexität der heutigen Arbeitswelt nicht mehr wegzudenken (Bell & Kozlowski, 2008). Die immer stärker werdende Orientierung von Organisationen, weg von starren hierarchischen Strukturen, hin zu agilen und anpassungsfähigen Teams, wurde im Global Human Capital Trends Bericht des Unternehmensberatungsnetzwerks Deloitte zu einem der drei wichtigsten Trends im Organisationskontext gezählt, in Deutschland und weltweit (McDowell, Agarwal, Miller, Okamoto, & Page, 2016). Darüber hinaus nehmen Teams aber auch in der verhaltensorientierten Organisationsforschung mit fast 20% Publikationsanteil (Humphrey & Aime, 2014) einen zentralen Platz ein. Aus diesen Gründen werden in dieser Arbeit primär Teams betrachtet.

Teams werden, angelehnt an Kozlowski und Ilgen (2006), definiert als (a) zwei oder mehr Individuen, die (b) für die Organisation relevante Aufgaben erfüllen, (c) mindestens ein gemeinsames Ziel teilen, (d) sozial interagieren, (e) in ihren Handlungen, Zielen, und Ergebnissen voneinander abhängig sind, und (f) in einen organisationalen Kontext eingebettet sind, der sie in ihren Handlungen begrenzt und gleichzeitig den Austausch mit anderen Einheiten der Organisation beeinflusst und regelt.

Aufbauend auf der Relevanz von Teams, entwickelte (West, 1996, 2000) einen der am häufigsten untersuchten, team-orientierten Ansätze im Reflexionsbereich. West verwendete dabei den Begriff *Team Reflexivität*, den er als „the extent to which team members collectively reflect upon the team's objectives, strategies, and processes, as well as their wider organizations and environments, and adapt them accordingly" (S. 3) definierte. Im Gegensatz zu anderen Reflexionsansätzen (z.B. Anseel, Lievens, & Schollaert, 2009), hebt West den Begriff *Reflexivität* über den Reflexionsprozess hinaus und integriert nicht nur die Erkenntnis durch Selbstbeobachtung („turning back on the self", West, 1996, S. 560), sondern auch die

Planung und Implementation von Folgehandlungen, die auf den durch Reflexion gewonnenen Einsichten basieren.

Zentrale Annahme von Wests Reflexivitätskonzepts ist, dass sich aus wiederholten Zyklen des Reflektierens, Planens, und Implementierens die Leistung von Teams verbessert (siehe Abbildung 1.1). Reflexive Teams seien nicht-reflexiven Teams dadurch überlegen, weil sie ‚bewusst' handeln. Ein nicht-reflexives Team zeichnet sich laut West (2000) hingegen durch eine Tendenz aus, lediglich auf die derzeitigen Gegebenheiten instinktiv zu reagieren, was von West auch als Handlung ohne Bewusstheit der Handlung bezeichnet wird.

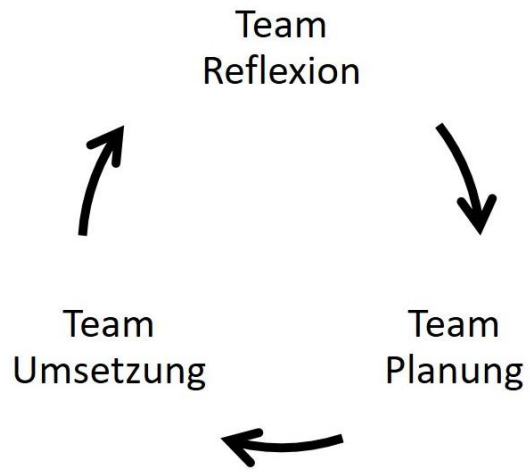

Abbildung 1.1: Team Reflexivitätszyklus angelehnt an West (2000).

1.2 Zielsetzung dieser Dissertation

Wests (1996, 2000) theoretische Beiträge inspirierten eine Vielzahl empirischer Folgearbeiten (z.B. Schippers, Hartog, Koopman, & Wienk, 2003; De Dreu, 2007; De Jong & Elfring, 2010; Wiedow & Konradt, 2011), die sich primär mit dem Zusammenhang von Team Reflexivität und Team Leistung, bzw. Team Innovation auseinandersetzten (siehe Konradt, Otte, Schippers, & Steenfatt, 2016, für ein Review). Anzumerken ist allerdings, dass in einer Mehrheit der Fälle nicht das übergeordnete Konstrukt der Team Reflexivität untersucht wurde, sondern der in die Team Reflexivität integrierte Prozess der Team Reflexion.

Tabelle 1.1: Übersicht über bisherige Ergebnisse aus der Teamreflexivitätsforschung

Authors	Design	Sample	Input	Process	Moderator	Output
Brav et al. (2009)	Cross-sectional, field	31 work teams		Team reflection		Self-organizational activities (+)
Carter & West (1998)	Cross-sectional, field	18 teams		Team reflection		Audience appreciation (0) Manager's ratings (+) Job feelings (0) Affective well-being (+) General well-being (+)
Dayan & Basir (2009)	Cross-sectional, field	107 product development teams	Transactive Memory System (+) Goal Clarity (+) Team Empowerment (+) Interactional Justice (+)	Team reflection		Product success (+)
De Dreu (2002)	Cross-sectional, field	32 organizational teams		Team reflection	Minority dissent (+)	Team innovation (0) Team effectiveness (-)
De Dreu (2007)	Cross-sectional, field	46 cross-functional teams		Team reflection	Cooperative outcome interdependence (+)	Team effectiveness (+) Learning (+) Information sharing (+)
De Jong & Elfering (2010)	Cross-sectional, field	73 teams from dutch tax department	Trust (+)	Team reflection		Team performance (0)
Facchin et al. (2006)	Cross-sectional, field	39 work teams		Team reflection		Team performance (+)
Fay et al. (2006)	Cross-sectional, field	46 health care teams			Team reflection and team climate inventory (+/-)	Innovation quality (+) Innovation quantity (0)
Gabelica et al. (2014)	Experimental	32 student teams		Team reflection		Performance (+)
Gabelica et al. (2014)	Experimental	106 student teams	Feedback (+)	Team reflection		Performance (+)
Gevers et al. (2001)	Longitudinal	22 student project teams		Team reflection		Meeting deadlines (0)
Gevers et al. (2009)	Longitudinal, field	48 student teams		Temporal reflection		Supervisor-rated performance (0)

Authors	Design	Sample	Input	Process	Moderator	Output
Gurtner et al. (2007)	Experimental	49 student teams		Team reflection		Team performance (0)
Gurtner (2009)	Cross-sectional, field	220 work teams	social stressors (0) interdependence (+) goal clarity (+)	Team reflection	Task-interdependence (+)	Self-rated performance (+) Supervisor-rated performance (0)
Hirst & Mann (2004)	Longitudinal and field	56 R&D teams		Team reflection		Performance self (+) Performance customer (0) Performance manager (0)
Hirst et al. (2004)	Longitudinal and field	50 R&D teams	facilitative leadership (+)	Team reflection		Team leader rated team performance (+)
Hoegl & Parboteeah (2006)	Cross-sectional, field	145 software development teams	project type (+) social skills (+) project management skills (+)	Team reflection		Self-rated team effectiveness (+) Self-rated efficiency (0)
Konradt & Eckardt (2016)	Longitudinal and field	97 student teams	feedback (0)	Team reflection		Team performance (+)
Konradt et al. (2015)	Experimental	98 student teams	feedback (+)	Team reflection		Team performance (+)
Lee & Sukoco (2011)	Cross-sectional, field	77 product development teams		Team reflection		Unlearning (+) Product innovation (+)
MacCurtain et al. (2010)	Cross-sectional, field	39 top management teams	trustworthiness (+)	Team reflection		Market innovation (+)
Nederveen Pieterse (2011)	Experimental	48 student teams	LGO (0) PPGO (0) gender diversity (+)	Team reflection		Team performance (0)
Oertel & Antoni (2015)	Longitudinal and field	37 student teams	TMS (+)	Team reflection		Transactive memory systems (0)
Savelsbergh et al. (2009)	Cross-sectional, field	19 operational teams		Outcome reflection Process reflection		Team performance (0) Team performance (0)
Schippers et al. (2003)	Cross-sectional, field	54 work teams	outcome inter-dependence (+) group longevity (+)	Team reflection		Self-rated performance (+)

Authors	Design	Sample	Input	Process	Moderator	Output
Schippers et al. (2008)	Cross-sectional, field	32 work teams	shared vision (+) transformational leadership (+)	Team reflection		Team performance (+)
Schippers et al. (2013)	Longitudinal	73 student teams		Team reflection		Team performance (+)
Schippers et al. (2015)	Cross-sectional, field	98 primary health care teams		Team reflection	work load (+) quality of environment (-)	Team innovation (+)
Shin et al. (2016)	Longitudinal, field	108 work teams	team openness to experience (+) team LGO (+) team PPGO (+)	Team reflection		Team creative performance (+)
Tjosvold et al. (2003)	Cross-sectional, Field	100 work teams	cooperative team approach (+) competitive team approach (-)	Team reflection		In-role performance (+) Identification (+)
Tjosvold et al. (2004)	Cross-sectional, Field	100 work teams	cooperation (+) competition (-)	Team reflection		Innovation (+)
van Ginkel & van Knippenberg (2009)	Experimental	125 student teams	Team Reflection (+)	Task representation		Decision quality (+)
Van Ginkel et al. (2009)	Experimental	84 student teams	Team reflection (+)	Task representation		Decision quality (+)
Wiedow & Konradt (2011)	Cross-sectional, field	50 work teams		Team reflection		Self-rated performance (+) Other-rated performance (-)

Anmerkung. Um Unschärfe durch die Übersetzung der englischen Begriffe zu vermeiden, wurde die Tabelle in englischer Sprache erstellt. Es wurden nur Studien mit tatsächlichen Teams integriert. Das mathematische Symbol in der Input-Spalte gibt die Richtung der Beziehung zum nachfolgenden Prozess an. Das Symbol hinter der Moderator-Spalte gibt die Richtung der Interaktion mit diesem Prozess auf den nachfolgenden Output an. Das mathematische Symbol hinter der mit Output bezeichneten Spalte gibt schließlich die Richtung der Beziehung zwischen Prozess und Output an.

Dieses Vorgehen wird auch in der vorliegenden Arbeit verfolgt, was der Tatsache geschuldet ist, dass überraschenderweise bereits auf der Ebene des Reflexionsprozesses eine ausgesprochene Unklarheit hinsichtlich der Frage herrscht, wie Team Reflexion und Team Leistung zusammenhängen. Entgegen der ursprünglich erwarteten positiven Zusammenhänge, offenbart sich nämlich bei einer genauen Betrachtung der bisherigen Forschung (Stand Februar 2017) ein ausgesprochen heterogenes Bild (siehe Tabelle 1.1). Von insgesamt 34 Untersuchungen im Team Reflexivitätskontext, zeigte sich lediglich bei 15 von 30 (50%) untersuchten Beziehungen zu Leistung eine positive Relation zwischen Team Reflexion und Leistung. Diese Zahl sinkt weiter, wenn man jene Studien ausschließt, die subjektive Operationalisierungen der Ergebniskriterien eingesetzt haben, wie zum Beispiel selbst-bewertete Leistung. Moreland und McMinn (2010) resümierten darüber hinaus, dass die Aussagekraft mehrere Reflexivitätsstudien durch vielfältige Mängel geschmälert wird, wie zum Beispiel die primäre Verwendung von Querschnittsdesigns, kleine Stichprobengrößen, oder defiziente Messmethoden.

Insgesamt kann die Ergebnislage daher so interpretiert werden, dass wichtige Elemente des Zusammenhangs zwischen Team Reflexion und Leistung bislang noch nicht ausreichend verstanden sind. Dieses unvollständige Verständnis ist von einem besonderen Interesse, weil in einer Vielzahl von Arbeiten für die Relevanz des Team Reflexionsprozesses im organisationalen Kontext argumentiert wird (z.B. Le Pine, Piccolo, Jackson, Mathieu, & Saul, 2008; Zajac, Gregory, Bedwell, Kramer, & Salas, 2014; Schippers, Edmondson, & West, 2014), und andere Autoren sogar argumentieren, dass Unternehmen die Leistung ihrer Mitarbeiter um 20 bis 25 Prozent steigern können, wenn sie reflektieren (Tannenbaum & Cerasoli, 2013). Es ist also anzunehmen, dass eine Aufklärung der heterogenen Befundlage nicht nur einen wichtigen Beitrag zur Teamforschung leistet, sondern auch darüber hinaus aktiv zur Gestaltung von Personalentwicklungs- unter Interventionsmaßnahmen beiträgt.

Aus diesen Überlegungen ergibt sich für die vorliegende Dissertation das übergeordnete Ziel, mögliche Ursachen für die beschriebene Heterogenität der Beziehung zwischen Team Reflexion und Leistung zu identifizieren. Diesem Ziel wird sich dabei theoretisch und empirisch in drei aufeinander aufbauenden Studien genähert.

Die erste Studie setzt sich mit dem Problem der Konzeption und Messung von Team Reflexivität auseinander. Über vier einzelne Untersuchungen mit insgesamt 803 Studierenden und Arbeitnehmern hinweg, wird ausführlich die Entwicklung und Validierung einer neuen Skala zur fragebogenbasierten Erfassung des Kern-Konstrukts der Team Reflexion dargelegt (siehe auch Kapitel 1.2).

Darauf aufbauend wird in Studie 2 mit der neu entwickelten Skala die Beziehung zwischen Team Reflexion und Leistung unter Berücksichtigung von wichtigen zeitlichen Aspekten untersucht. In einer Laborstudie mit 47 drei-personen Teams liegt dabei der Fokus auf folgenden Fragen: Erstens, inwiefern wird Team Reflexion gemäß selbst-regulatorischer Annahmen (Carver & Scheier, 1998; Konradt et al., 2016) im Kontext von longitudinalen Modellen von vorangehender Leistung beeinflusst? Zweitens, welche Effekte haben qualitative und quantitative Team Reflexion auf die Leistungsverbesserung, und drittens, welche Rolle nimmt dabei die Implementation von Lösungsansätzen ein? Als vierten Punkt wird schließlich untersucht, inwiefern selbst- und fremdeingeschätzte Team Reflexion zusammenhängen (siehe auch Kapitel 1.3).

In der dritten Studie wird die Frage nach der zeitlichen Komponente als wichtiger Einflussfaktor auf die Beziehung zwischen Team Reflexion und Leistung tiefergehender untersucht. Dabei wird auf der Punctuated Equilibrium Theorie (Gersick, 1991; Tushman & Romanelli, 1985) aufgebaut, die um relevante, teamtheoretische Elemente erweitert wird. Anschließend wird in einem longitudinalen Setting mit sechs Messzeitpunkten untersucht, inwiefern die in Studie 1 entwickelten Dimensionen des Reflexionsprozesses mit Leistung

zusammenhängen, wenn man sie im Kontext einer dynamischen, organisationalen Umwelt betrachtet (siehe auch Kapitel 1.4).

1.3 Beitrag Studie 1

Im 2016 erschienenen Review zur Team Reflexivität von Konradt et al. wurden mehrere konzeptionelle und messmethodische Probleme von Wests (2000) dargelegten Reflexivitäts-Konzepts identifiziert und analysiert, die möglicherweise zur beschriebenen Heterogenität der Ergebnisse in der Reflexivitätsforschung beigetragen haben. Ein zentrales Element dieser Analyse ist die Messung von Team Reflexivität. Eine genaue Betrachtung der am häufigsten (Konradt et al., 2016) verwendeten Skala von Carter und West (1998) und darauf aufbauenden Abwandlungen, offenbart hierbei mehrere Probleme hinsichtlich der Inhaltsvalidität und der Konstruktvalidität.

Grundsätzlich ist festzuhalten, dass die Skala nach Carter und West (siehe Tabelle 1.2) den Ansatz verfolgte, das gesamte Konzept Team Reflexivität abzubilden. Eine genauere Betrachtung zeigt jedoch, dass eher der der Reflexivität untergeordnete Reflexionsprozess, als der Planungs- und Implementationsprozess, in den Items abgebildet wird. West (2000) hält zur Definition von Team Reflexion allerdings fest, dass sie aus „attention, awareness, monitoring, and evaluation of the object of reflection" (p. 4) besteht, und Verhaltensweisen wie (1) Infrage stellen, (2) Planen, (3) exploratives Lernen, (4) Analysieren, (5) Explorieren, oder (6) lernen auf einem Meta-Level beinhaltet. Tatsächlich konzentriert sich die Skala hauptsächlich auf das Diskutieren von Prozessen (Schippers, Den Hartog, & Koopman, 2007), weshalb die tatsächliche Abbildung der genannten Verhaltensweisen in Zweifel gezogen werden kann. Darüber hinaus werden die Team Prozesse Planen und Implementieren höchstens implizit über die Items 4 und 5 erfasst.

Tabelle 1.2: Team Reflexivität gemessen mit der Skala nach Carter und West (1998)

Nr.	Items	Tatsächliches Konstrukt
1	Wir *überprüfen* regelmäßig unsere Ziele.	Monitoring
2	Wir *diskutieren* häufig über die Methoden, wie wir unsere Arbeit machen.	Diskutieren von Prozessen
3	Wir *diskutieren* regelmäßig, ob wir effektiv zusammenarbeiten.	Diskutieren von Prozessen
4	Wenn sich die Umstände ändern, *passen* wir auch unsere Ziele den Veränderungen an.	Team Anpassungsfähigkeit
5	Wir *passen* unsere Strategien häufig an.	Team Anpassungsfähigkeit
6	Wir *sprechen* oft darüber, ob wir angemessen miteinander kommunizieren.	Diskutieren von Prozessen
7	Wir *diskutieren* regelmäßig darüber, ob die Art und Weise, wie wir unsere Arbeit tun, angemessen ist.	Diskutieren von Prozessen
8	Die Art und Weise, wie wir zu Entscheidungen kommen, wird regelmäßig *hinterfragt* und verändert.	Diskutieren von Prozessen & Team Anpassungsfähigkeit

Anmerkung. Linke Spalte: Team Reflexivitätsskala durch van Dick & West (2005) ins Deutsche übersetzt. Rechte Spalte: Zugeordnete Konstrukte.

Zusammengenommen kann die Inhaltsvalidität der Skala also als nicht ausreichend bewertet werden. Hinsichtlich der Konstruktvalidität stellt sich zuerst die Frage, ob eine derartig diverse Konzeption in Form eines sogenannten Omnibuskonstrukts (Reflexion, Planen, und Implementation) überhaupt sinnvoll ist, da dies mit verschiedenen Problemen einhergeht (Konradt et al., 2016). Zuerst erlaubt der Ansatz keine differenzierte Betrachtung der einzelnen untergeordneten Prozesse der Team Reflexivität, wodurch eine exaktere Diagnose von Abläufen im Team verhindert wird. Darüber hinaus unterliegt der Ansatz aber auch einer generellen ‚je mehr desto besser' Einschränkung. Dabei wird der Annahme gefolgt, dass mehr Reflexion, mehr Planung, und mehr Implementation im Sinne einer ‚One size fits all' Lösung, grundsätzlich zu einer besseren Leistung führt, was sich durch die

bisherige Forschung allerdings nicht belegen ließ (Konradt et al., 2016). Drittens lässt sich diese Konzeptualisierung als sogenanntes ‚Compositions Modell' bezeichnen (Kozlowski & Klein, 2000). Derartige Modelle gehen davon aus, dass die Aggregation mehrere Variablen zu einer einzigen Variable auf einem höheren Level über einen einfachen Mittelwert abzubilden ist. Dabei wird angenommen, dass die untergeordneten Prozesse ähnlich gewichtet sind, und somit im gleichen Verhältnis zueinanderstehen, als auch zu den Ergebnisvariablen wie Leistung (Mathieu, Tannenbaum, Donsbach, & Alliger, 2014). Da das Verhältnis von Reflexion zu Planung und zu Implementation hingegen kaum untersucht wurde, kann auch diese Annahme in Zweifel gezogen werden.

Ebenfalls negativ auf die Validität der Carter und West Skala wirken sich darüber hinaus Konfundierungen mit Reflexivitäts-nahen Konstrukten aus. Obgleich West argumentiert, dass Team Monitoring (Definiert als Beobachtung von Handlungen von Teammitgliedern, sowie des Ausschauhaltens nach Fehlern oder Diskrepanzen zwischen Ist- und Soll-Zuständen, Marks, Mathieu, & Zaccaro, 2001) zur Reflexion dazugehört, wird es in einem Großteil der organisationsbezogenen Forschung (z.B. De Jong & Elfring, 2010) als ein eigenständiges Konstrukt erfasst. Gleiches gilt für das Konstrukt Team Adaptation (Definiert als das Anpassen von Verhaltensweisen an Anforderungen aus der Umwelt, Pulakos, Arad, Donovan, & Plamondon, 2000), welches ebenfalls ein etabliertes und eigenständiges Konstrukt in der OB Forschung darstellt (für ein Review siehe Baard, Rench, & Kozlowski, 2014). Beide Konstrukte sollten daher gesondert behandelt werden.

Zusammenfassend ist festzuhalten, dass die bisherige Messung von Team Reflexivität/Reflexion durch eine mangelhafte Inhalts- und Konstruktvalidität beeinträchtigt wurde. Es folgt, dass relevante Varianz von Kriteriumsmaßen (z.B. Leistung) nicht aufgeklärt werden konnte, und dadurch wichtige Zusammenhänge möglicherweise verdeckt wurden (siehe Messick, 1995).

Insgesamt scheint es also notwendig, die Konzeptualisierung und Messung von Reflexion zu überdenken. In Studie 1 wird daher vorgeschlagen, den Begriff der Team Reflexivität nicht mehr auf ein einzelnes psychologisches Konstrukt zu beziehen, sondern als Bezeichnung für ein Framework von verschiedenen, mit einander verbundenen reflexiven Prozesse zu verwenden. Gleichzeitig sollte innerhalb des Frameworks jeder Prozess einzeln gemessen werden. Des Weiteren wird eine neue, mehrdimensionale Skala (REMINT – Reflection Measure for Individuals and Teams) zur Messung des Kern-Reflexivitätsprozesses Reflexion entwickelt. Dabei wird insbesondere auf die Feststellung von Konradt et al. (2016) eingegangen, dass bisherige Skalen zur Messung von Team Reflexion ausschließlich die Häufigkeit (Quantität) erfasst haben. Somit wurden allerdings wichtige qualitative Merkmale wie Tiefe und Detail, außeracht gelassen. Schippers, Edmondson, und West (2014) hielten ebenfalls fest, dass die Konzeptualisierung von Qualität ein wichtiger Schritt in der Reflexionsforschung wäre, und Moreland und McMinn (2010) diskutierten sogar, dass Qualität wichtiger sei als Quantität. Diese neue Skala wird nun in Studie 1 im Rahmen von vier einzelnen Untersuchungen an verschiedenen Stichproben (Studierende und Arbeitnehmer) ausführlich validiert, um nicht nur Belege für die Konstrukt-, sondern auch die Kriteriumsvalidität, und somit den Mehrwert der Neukonzeptualisierung zu erbringen.

1.4 Beitrag Studie 2

Eine häufig kritisierte Unzulänglichkeit innerhalb der Organisationsforschung ist die oftmals verwendete, statische Perspektive auf Individuen und Teams (Humphrey & Aime, 2014). Bei der Anwendung von Forschungsdesigns, die die zeitliche Dimension nicht berücksichtigen, entsteht die Gefahr relevante Zusammenhänge zu verdecken, oder auf fundamentaler Ebene falsch wiederzugeben. „Nur, wenn Mitarbeiter, Teams und Organisationen über einen längeren Zeitraum begleitet werden, können die Dynamiken in Organisationen abgebildet werden“ (Kauffeld, 2014, S. 9). Die Kritik an Untersuchungen

ohne Berücksichtigung der zeitlichen Dynamik ist nicht neu. Bereits in Starbuck (1965) wird darauf verwiesen, dass für eine aussagekräftige Analyse mindestens zwei Messzeitpunkte benötigt werden, optimaler Weise mehr. Nichtdestotrotz wird in Mathieu et al. (2014) immer noch kritisiert, dass Teamuntersuchungen selten longitudinale Messungen von Kriterien vornehmen. Mitunter wurden sogar die Vorteile von längsschnittlichen Untersuchungen verworfen, indem beispielsweise Daten mit zeitlicher Struktur mittels Differenzwertbildung zusammengefasst wurden (z.B. Barry & Stewart, 1997), wodurch relevante Varianz verloren geht (Dormann, Zapf, & Perels, 2010).

Zusätzlich zum Informationsverlust führt eine unzureichende Auseinandersetzung mit längsschnittlichen Strukturen allerdings auch schnell zu falschen Vorstellungen von Zusammenhängen. In der Reflexivitätsforschung ist hier vor allem eine Annahme hinsichtlich der Konstanz von Zusammenhängen zwischen beispielsweise einem Prozess und einem Outcome zu kritisieren. Ausgehend von der zuvor beschriebenen ‚je mehr desto besser' Perspektive, wird häufig davon ausgegangen, dass sich ein bestimmter Prozess immer in gleicher Weise auf ein Kriterium auswirkt. Da die organisationale Umwelt aber dynamisch ist (Bell & Kozlowski, 2008; Humphrey & Aime, 2014), muss ein positiver Effekt zu einem gewissen Zeitpunkt nicht bedeuten, dass dieser Effekt zu einem späteren Zeitpunkt gleichermaßen ausfällt. Stattdessen kann der Effekt von Drittvariablen abhängen, die ihrerseits ebenfalls variieren können. Hollenbeck, Moon, Ellis, West, Ilgen, Sheppard, et al. (2002) zeigten beispielsweise in einer Untersuchung mit 80 Viererteams, dass die durch hohe kognitive Leistungsfähigkeit bestehenden Vorteile eines Teams wieder negiert werden können, wenn die Struktur eines Teams nicht zu den Umweltbedingungen passt. Wenn aber bereits vermeintlich triviale Zusammenhänge, wie kognitive Fähigkeiten und Leistung durch eine einfache Manipulation der Teamstruktur durcheinandergebracht werden können,

verlangen über die Zeit verlaufende Zusammenhänge eine noch gewissenhaftere Auseinandersetzung.

Zur Berücksichtigung von Daten mit einer longitudinalen Struktur müssen neben fortgeschrittenen statistischen Analysemethoden allerdings auch weiterentwickelte theoretische Modelle vorausgesetzt werden. Ein Modell, das im Zentrum dieser Arbeit steht, ist das *Recurring Phase Model of Team Processes* von Marks et al. (2001). Eine der zentralen Funktionen des Modells ist die zeitliche Ordnung von Teamprozessen. Team Prozesse werden dabei als Handlungen von Teammitgliedern bezeichnet, die mittels kognitiver, verbaler und behavioraler Aktivitäten Inputs (zum Beispiel Feedback zu einer vergangenen Leistung) zu Outputs konvertieren, um die Aufgabenbearbeitung zu organisieren und die Zielerreichung des Teams zu unterstützen (Marks et al., 2001).

Marks und Kollegen betrachten Team Prozesse allerdings nicht mehr im isolierten Kontext eines einzelnen ‚Input-Prozess-Output' Modells (IPO, McGrath, 1964), sondern im Rahmen von wiederkehrenden Zyklen, sogenannten *Team Performance Episoden*. Jede dieser Episoden besteht aus einem eigenen Input-Prozess-Output Element, wobei der Output, das heißt das Ende eines Zyklus', zum Input des nächsten Zyklus' wird (siehe Abbildung 1.2).

Marks et al. unterscheiden dabei zwischen zwei verschiedenen Episodentypen. Transitionsepisoden bezeichnen Episoden, in denen Teams primär damit beschäftigt sind vorangehende Handlungen zu evaluieren (bspw. Team Reflexion) und neue zu planen.

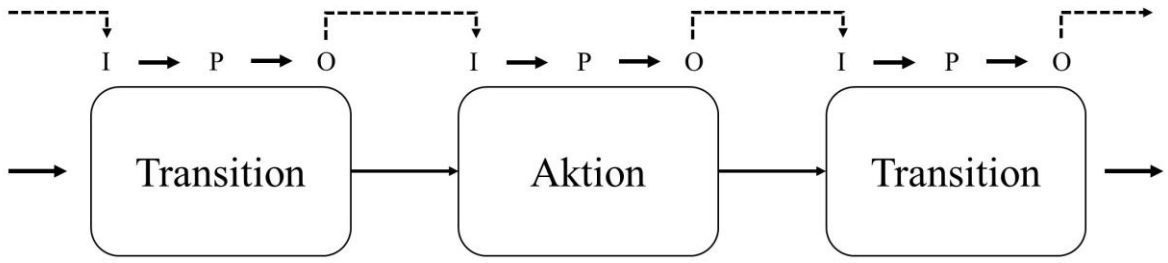

Abbildung 1.2: Recurring Phase Model of Team Processes nach Marks et al. (2001)

Anmerkung. I: Input. P: Prozess. O: Output.

Aktionsepisoden hingegen bezeichnen Episoden, in denen Teams sich direkt mittels aktiver Handlungen mit der Zielerreichung auseinandersetzen (bspw. Implementation). Der Vorteil des Recurring Phase Models ist somit die explizite Berücksichtigung von Parametern aus vorangegangenen Episoden (bspw. vorherige Leistung), wodurch sich longitudinale Zusammenhänge besser beschreiben und erklären lassen.

In der Team Reflexivitätsforschung wurden longitudinale Zusammenhänge und die zeitliche Veränderung von beispielsweise Reflexion lange nicht betrachtet, was möglicherweise auf der Annahme beruhte, dass Team Reflexion unter allen Umständen positiv mit Leistung zusammenhängen sollte. Erst Schippers, Homan, und van Knippenberg (2013) zeigten in einer Untersuchung mit 73 Studierendenteams die über zwei Messzeitpunkte an ihrer Bachelorthesis arbeiteten, dass die angenommene Konstante Beziehung zwischen Reflexion und Leistung die tatsächlichen Zusammenhänge nur unzureichend wiedergibt. Schippers et al. legten dar, dass Team Reflexion nur dann einen positiven Effekt auf die Team Leistung hatte, wenn die Leistung zu einem vorherigen Messzeitpunkt niedrig war. Teams, die bereits zum vorherigen Messzeitpunkt eine gute Leistung aufwiesen, und anschließend viel reflektierten, zeigten hingegen keinen weiteren Leistungsgewinn gegenüber wenig reflektierenden Teams.

Dieser Befund deckt sich mit der Konzeption von Team Reflexion (und Planen und Implementation gleichermaßen) als Diskrepanz reduzierender Prozess (Konradt et al., 2016). Aufbauend auf der Selbst-Regulationstheorie (Carver & Scheier, 1998) und der Team-Regulationstheorie von DeShon, Kozlowski, Schmidt, Milner, und Wiechmann (2004) wird dabei angenommen, dass Teams ihr Verhalten selbstständig regulieren um ihre Ziele zu erreichen. Wichtiger Bestandteil der Selbst-Regulation sind Abweichungen zwischen dem derzeitigen Grad der Zielerreichung, und dem gewünschten Grad der Zielerreichung. Große Diskrepanzen zwischen beiden Zuständen initiieren im Team daher stärker ausgeprägte

Regulationsprozesse (z.B. Reflexion und Implementation), wohingegen kleine Diskrepanzen die selbst-regulatorischen Anstrengungen eines Teams solange verringern, bis die Diskrepanz wieder hinreichend groß ist.

Ein wichtiger Diskrepanz-Indikator ist Feedback (Gabelica, Van den Bossche, De Maeyer, Segers, & Gijselaers, 2014; Konradt, Schippers, Garbers, & Steenfatt, 2015). Zwar sagt die Selbst-Regulationstheorie auch vorher, dass Individuen und Teams aus intrinsischer Motivation heraus bewusst Diskrepanzen schaffen können (Carver & Scheier, 1998), dies soll aber nicht im Fokus dieser Arbeit stehen. Stattdessen wird davon ausgegangen, dass vorangehende Team Performance Episoden ein wichtiger selbst-regulatorischer Indikator in Form von Feedback, beispielsweise durch vorangegangene Leistung, für das Team sind (Konradt et al., 2016).

Konradt und Kollegen (2016) argumentierten, dass neben der im vorherigen Abschnitt erläuterten Messproblematik, auch die kaum vorhandene Berücksichtigung der zeitlichen Dimension für die heterogene Befundlage in der Reflexivitätsforschung verantwortlich sein kann, weil dadurch vorangegangene Performance Episoden nicht berücksichtigt wurden. Vorangegangene Episoden steuern aber zu einem wichtigen Teil das Verhalten von Teams (Marks et al., 2001). Von diesem Standpunkt aus wäre es beispielsweise nicht verwunderlich, wenn ein Team mit herausragender Leistung weniger reflektiert, weil es zu einem vorangegangenen Zeitpunkt bereits gute Leistung gezeigt hat. Ein Beleg für diese Annahme wurde bisher allerdings erst indirekt in Konradt und Eckardt (2016) gefunden, die zeigen konnten, dass bei über vier Messzeitpunkten ansteigende Leistung mit sinkender Team Reflexion zusammenhängt. Ein direkter Effekt von vorangegangener Leistung auf Reflexion konnte bisher noch nicht bestätigt werden, weshalb das erste Ziel der zweiten Studie das Testen dieser Annahme unter Laborbedingungen ist. Dabei wird zur Messung von Team Reflexion erstmals die in Studie 1 konstruierte Skala angewandt, um zusätzlich zu

untersuchen, in welcher Beziehung Qualität und Quantität von Team Reflexion im Rahmen einer zeitlich begrenzten Reflexionsphase einerseits zum angegliederten Aktionsprozess der Implementation, andererseits aber auch zur Teamleistung stehen.

Schlussendlich geht Studie 2 auch zum ersten Mal auf einen Vergleich von verschiedenen Messmethoden von Teamreflexion ein. Mit Ausnahme von zwei einzelnen Untersuchungen (Gabelica, Van den Bossche, De Maeyer, & Gijselaers, 2014; Gurtner, Tschan, Semmer, & Nägele, 2007), hat sich die gesamte Teamforschung bislang ganz auf die Verwendung von Fragebögen konzentriert. Fragebögen können allerdings diversen Urteilsverzerrungen (Biases) unterliegen, wie zum Beispiel Halo oder Recency Effekten (Podsakoff & Organ, 1986). Bisher hat noch keine Studie im Reflexionskontext untersucht, inwiefern die aus Fragebögen gewonnenen Daten tatsächlich durch Urteilsverzerrungen beeinträchtigt werden. Ein Bias ist besonders hinsichtlich der Unterscheidung von quantitativer und qualitativen Reflexion interessant. Es ist anzunehmen, dass die Einschätzung wie oft ein Verhalten auftritt (d.h. die Quantität) für Probanden einfacher zu bewältigen ist, weil es sich um die schlichte Häufigkeit des Auftretens handelt. Die korrekte Einschätzung der Qualität eines Verhaltens sollte hingegen ein umfassenderes Wissen voraussetzen, weil dafür ein Verständnis von Qualität und damit verbundenen Kriterien einhergeht. Daher wird in Studie 2, zusätzlich zu den in Studie 1 entwickelten Skala, ein verhaltensbasiertes Maß zur Erfassung der Team Reflexion konzipiert und eingesetzt.

Für die Studie wird auf das bereits erprobte Reisebüro-Paradigma zurückgegriffen (z.B. Wiedow, Konradt, Ellwart, & Steenfatt, 2013), das allerdings in vielerlei Hinsicht weiterentwickelt wird, um die Komplexität zu steigern und eine größere Vielfalt an möglichen Strategien zu bieten. Teams müssen dabei über zwei Runden die Aufgabe eines fiktiven Reisebüros übernehmen und Kundenanfragen zu möglichen Urlaubszielen bearbeiten. Die Aufgabe wird dadurch erschwert, dass die notwendigen Informationen im Stil

eines Hidden Profiles (Stasser & Titus, 1985) asymmetrisch über die Teammitglieder verteilt sind. Sie sind somit einerseits gezwungen, Lösungsstrategien für die eigentliche Aufgabe zu entwickeln, und gleichzeitig ihre Kommunikation zu optimieren.

1.5 Beitrag Studie 3

Die dritte Studie setzt sich schließlich mit dem Team Reflexionsprozess in einem größeren longitudinalen Kontext auseinander, um das Verständnis von der Dynamik des Zusammenhangs zwischen Team Reflexion und Leistung weiter zu vertiefen. Aussagekräftige Team Reflexionsforschung basierte bisher meist auf kontrollierten Laborbedingungen (siehe z.B. Studie 2). Dabei kann jedoch die Dynamik der organisationalen Realität meist nur begrenzt berücksichtigt werden. Eine wichtige Theorie zur Beschreibung dieser Dynamik ist die *Punctuated Equilibirum Theorie* (Theorie des punktierten Gleichgewichts; Gersick, 1991; Tushman & Romanelli, 1985). Aufbauend auf evolutionstheoretischen Annahmen, wird davon ausgegangen, dass sich Teams (und Organisationen[1]) immer in Richtung eines Equilibriums, das heißt eines stabilen, selbsterhaltenden Zustands entwickeln. *Equilibriumsperioden* werden allerdings von wiederkehrenden *Punktuationen* unterbrochen. Eine Punktuation läutet damit eine kürzere, *revolutionäre Periode* ein, in denen Teams und Unternehmen sich radikal und fundamental ändern und an die neu entstandenen Begebenheiten anpassen müssen. Revolutionäre Perioden werden durch interne oder externe Schocks ausgelöst, wie zum Beispiel Änderungen in der Führung (bspw. Austausch des Top-Managements), fundamentale technologische Umwälzungen (bspw. neue Kommunikationstechnologien), oder Wirtschaftskrisen (Romanelli & Tushman, 1994; Tushman & Rosenkopf, 1992; Zellmer-Bruhn, 2003).

Überraschenderweise hat sich die empirische Forschung trotz der Prominenz des Modells (Humphrey & Aime, 2014) hauptsächlich mit der Frage auseinandergesetzt, ob die

[1] Zur Vereinfachung wird im Folgenden nur von Teams gesprochen, allerdings kann das Modell ebenfalls auf Organisationen angewandt werden.

Entwicklung eines Teams vom Beginn bis zum Ende eines einzelnen Projekts dem Rhythmus der von der Theorie prognostizierten Änderungstypen entspricht (z.B. Gersick, 1989; Okhuysen & Waller, 2002). Kaum eine Studie wurde hingegen mit dem Ziel durchgeführt, die tatsächlichen Anpassungsprozesse von Teams im Kontext der Theorie zu untersuchen. Eine mögliche Ursache hierfür könnte sein, dass die Theorie lediglich auf die *Struktur* von Veränderungen ausgerichtet ist, das heißt, in welche Kategorien Veränderungen eingeteilt werden können (Equilibriums- und Revolutionsperioden). Es fehlt allerdings eine theoretische Grundlage, die Prozesse in diese Dynamik integriert, damit wiederum prüfbare Hypothesen abgeleitet werden könnten.

Der erste Beitrag dieser Studie ist daher theoretischer Natur und erweitert die Punctuated Equilibrium Theorie um eine Prozesskomponente. Aufbauend auf den Annahmen von Marks und Kollegen (2001) wird die Integration miteinander verbundener Performance Episoden vorgeschlagen. Darüber hinaus wird argumentiert, dass die Effekte von Team Prozessen auf Ergebniskriterien in Equilibriums- und revolutionären Perioden von den jeweiligen, perioden-spezifischen Anforderungen abhängig sind. Speziell gehen die Anforderungen laut Gersick mit der Schaffung und dem Erhalt von sogenannten *tiefen Strukturen* einher (deep structures, Gersick, 1991).

Tiefe Strukturen können als die zugrundeliegende Ordnung eines Systems verstanden werden. Sie sind durch komplexe, verbundene Elemente gekennzeichnet, die in ihrem Zusammenwirken die Funktion und den Erhalt des Systems determinieren. Gersick (1991) beschrieb, dass während revolutionären Perioden die tiefe Struktur eines Teams aufgebaut werden muss. March (1991) bezeichnete die damit einhergehenden Aktivitäten auch als *Explorationsprozesse*, da sie auf die Entdeckung von etwas Neuem ausgerichtet sind. Hingegen muss in Equilibriumsperioden die tiefe Struktur gepflegt/gewartet (maintained) werden, was wiederum die Aufgabe von *Exploitationsprozessen* ist (vgl. March, 1991), also

die Nutzung bestehender Strukturen ohne ihre grundlegende Änderung und mit einem eher geringen Aufwand.

Aufbauend auf diesem theoretischen Gerüst wird qualitative Team Reflexion als Explorationsprozess konzipiert, und quantitative Team Reflexion als Exploitationsprozess. Dabei wird davon ausgegangen, dass für positive Effekte von Team Reflexion auf Leistung eine Passung zwischen Prozess (qualitativ und quantitativ) und kontextueller Anforderung (Revolution/Equilibrium) bestehen muss. Fehlende Passung kann hingegen zu keinen, oder sogar einem negativen Effekt von Team Reflexion auf Leistung führen.

Der zweite wichtige Beitrag von Studie 3 ist die empirische Testung dieser Annahmen. Dies geschieht im Rahmen eines Wirtschaftsplanspiels mit 86 Teams, die über insgesamt sechs Messzeitpunkte ein fiktives Unternehmen führen und dabei eine Vielzahl von Entscheidungen (Forschung und Entwicklung, Vertrieb, Personal, Marketing, etc.) treffen müssen. Der Unternehmenserfolg ist allerdings nicht nur von der Qualität der Entscheidungen abhängig, sondern auch vom Verhalten anderer fiktiven Unternehmen, die miteinander in Konkurrenz stehen. Auf diese Weise entsteht eine dynamische Umwelt für jedes Unternehmen, in der keine einzelne, optimale Lösung existiert. Als zusätzliche Besonderheit tritt im Planspiels nach dem dritten Messzeitpunkt eine Punctuation in Form einer Wirtschaftskrise auf, wodurch die Unternehmen gezwungen werden sich anzupassen. Neben der systematischen Analyse der Effekte von Team Reflexion auf Leistung, ist Studie 3 somit auch die erste Untersuchung, die explizit die Annahmen der Punctuated Equilibirum Theory anwendet, um die Effekte von organisationsbezogenen Verhaltensweisen zu erklären.

1.6 Quellen

Anseel, F., Lievens, F., & Schollaert, E. (2009). Reflection as a strategy to enhance task performance after feedback. *Organizational Behavior and Human Decision Processes, 110*, 23-35. doi:10.1016/j.obhdp.2009.05.003

Barry, B., & Stewart, G. L. (1997). Composition, process, and performance in self-managed groups: the role of personality. *Journal of Applied Psychology, 82*, 62-78. doi:10.1037/0021-9010.82.1.62

Bell, B. S., & Kozlowski, S. W. J. (2008). Active learning: Effects of core training design elements on self-regulatory processes, learning, and adaptability. *Journal of Applied Psychology, 93*, 296-316. doi:10.1037/0021-9010.93.2.296

Carter, S. M., & West, M. A. (1998). Reflexivity, effectiveness, and mental health in BBC-TV production teams. *Small Group Research, 29*, 583–601. doi:10.1177/1046496498295003

Carver, C. S., & Scheier, M. F. (1998). *On the Self-Regulation of Behavior*. New York, NY: Cambridge University Press.

De Dreu, C. K. W. (2002). Team innovation and effectiveness: The importance of minority dissent and reflexivity. *European Journal of Work and Organizational Psychology, 11*, 285–298. doi:10.1080/13594320244000175

De Dreu, C. K. W. (2007). Cooperative outcome interdependence, task reflexivity, and team effectiveness: a motivated information processing perspective. *Journal of Applied Psychology, 92*, 628-638. doi:10.1037/0021-9010.92.3.628

De Jong, B. A., & Elfring, T. (2010). How does trust affect the performance of ongoing teams? The mediating role of reflexivity, monitoring, and effort. *Academy of Management Journal, 53*, 535-549. doi:10.5465/AMJ.2010.51468649

DeShon, R. P., Kozlowski, S. W. J., Schmidt, A. M., Milner, K. R., & Wiechmann, D. (2004). A multiple goal, multilevel model of feedback effects on the regulation of individual and team performance. *Journal of Applied Psychology, 89*, 1035–1055. doi:10.1037/0021-9010.89.6.1035

Dormann, C., Zapf, D., & Perels, F. (2010). Quer-und Längsschnittstudien in der Arbeitspsychologie. In U. Kleinbeck, & K.-H. Schmidt (Hrsg.), *Enzyklopädie der Psychologie, Themenbereich D* (S. 923-1001). Göttingen: Hogrefe.

Dudenredaktion (2017, 25. Februar). Reflexion [Web log post]. Abgerufen unter http://www.duden.de/suchen/dudenonline/Reflexion

Fay, D., Borrill, C., & Amir, Z. (2006). Getting the most out of multidisciplinary teams: A multi-sample study of team innovation in health care. *Journal of Occupational and Organizational Psychology, 79*, 553–567. doi:10.1348/096317905£72128

Gabelica, C., Van den Bossche, P., De Maeyer, S., Segers, M., & Gijselaers, W. (2014). The effect of team feedback and guided reflexivity on team performance change. *Learning and Instruction, 34*, 86–96. doi:10.1016/j.learninstruc.2014.09.001

Gabelica, C., Van den Bossche, P., Segers, M., & Gijselaers, W. (2014). Dynamics of team reflexivity after feedback. *Frontline Learning Research, 2*, 64–91. doi:10.14786/flr.v2i2.79

Gersick, C. J. (1989). Marking time: Predictable transitions in task groups. *Academy of Management Journal*, *32*, 274-309. doi:10.2307/256363

Gersick, C. J. (1991). Revolutionary change theories: A multilevel exploration of the punctuated equilibrium paradigm. *Academy of Management Review, 16*, 10-36. doi:10.5465/AMR.1991.4278988

Gevers, J. M., van Eerde, W., & Rutte, C. G. (2001). Time pressure, potency, and progress in project groups. *European Journal of Work and Organizational Psychology, 10,* 205-221. doi:10.1080/13594320143000636

Gurtner, A. (2009). Reflexivity and social stressors in teams: An empirical study. *Leading Systems, 2,* 27-56.

Gurtner, A., Tschan, F., Semmer, N. K., & Nägele, C. (2007). Getting groups to develop good strategies: Effects of reflexivity interventions on team process, team performance, and shared mental models. *Organizational Behavior and Human Decision Processes, 102,* 127–142. doi:10.1016/j. obhdp.2006.05.002

Hirst, G., & Mann, L. (2004). A model of R&D leadership and team communication: The relationship with project performance. *R&D Management, 34,* 147-160. doi:10.1111/j.1467-9310.2004.00330.x

Hirst, G., Mann, L., Bain, P., Pirola-Merlo, A., & Richver, A. (2004). Learning to lead: The development and testing of a model of leadership learning. *The Leadership Quarterly, 15,* 311–327. doi:10.1016/j.leaqua.2004.02.011

Hoegl, M., & Parboteeah, K. P. (2006). Team reflexivity in innovative projects. *R&D Management, 36,* 113-125. doi:10.1111/j.1467-9310.2006.00420.x

Hollenbeck, J. R., Moon, H., Ellis, A. P., West, B. J., Ilgen, D. R., Sheppard, L., ... & Wagner III, J. A. (2002). Structural contingency theory and individual differences: examination of external and internal person-team fit. *Journal of Applied Psychology, 87,* 599-606. doi:10.1037/0021-9010.87.3.599

Humphrey, S. E., & Aime, F. (2014). Team microdynamics: Toward an organizing approach to teamwork. *The Academy of Management Annals, 8,* 443–503. doi:10.1080/19416520.2014.904140

Kauffeld, S. (2014). *Arbeits-, Organisations- und Personalpsychologie für Bachelor*. Berlin: Springer. doi:10.1007/978-3-642-42065-8_1

Konradt, U., & Eckardt, G. (2016). Short-term and long-term relations among reflection and performance in teams: Evidence from a four-wave longitudinal study. *European Journal of Work and Organizational Psychology*, *25*, 804–818. doi:10.1080/1359432X.2016.1160058

Konradt, U., Otte, K.-P., Schippers, M. C., & Steenfatt, C. (2016). Reflexivity in teams: A review and new perspectives. *The Journal of Psychology: Interdisciplinary and Applied, 150*, 153–174. doi:10.1080/00223980.2015.1050977

Konradt, U., Schippers, M. C., Garbers, Y., & Steenfatt, C. (2015). Effects of guided reflexivity and team feedback on team performance improvement: The role of team regulatory processes and cognitive emergent states. *European Journal of Work and Organizational Psychology*, *24*, 777–795. doi:10.1080/1359432X.2015.1005608

Kozlowski, S. W. J., & Ilgen, D. R. (2006). Enhancing the effectiveness of work groups and teams. *Psychological Science*, *7*, 77-124. doi:10.1111/j.1529-1006.2006.00030.x

LePine, J. A., Piccolo, R. F., Jackson, C. L., Mathieu, J. E., & Saul, J. R. (2008). A meta-analysis of teamwork processes: tests of a multidimensional model and relationships with team effectiveness criteria. *Personnel Psychology, 61*, 273-307. doi:10.1111/j.1744-6570.2008.00114.x

Lee, L. T.-S., & Sukoco, B. M. (2011). Reflexivity, stress, and unlearning in the new product development team: the moderating effect of procedural justice. *R&D Management*, *41*, 410-423. doi:10.1111/j.1467-9310.2011.00645.x

MacCurtain, S., Flood, P.C., Ramamoorthy, N., West, M.A., & Dawson, J.F. (2010). The top management team, reflexivity, knowledge sharing and new product performance: A

study of the Irish software industry. *Creativity and Innovation Management, 19*, 219–232. doi:10.1111/j.1467-8691.2010.00564.x

McDowell, T., Agarwal, D., Miller, D., Okamoto, T., & Page, T. (2016, 25. Februar). Organizational design: The rise of teams [Web log post]. Abgerufen unter: https://dupress.deloitte.com/dup-us-en/focus/human-capital-trends/2016/organizational-models-network-of-teams.html

March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science, 2*(1), 71-87. doi:10.1287/orsc.2.1.71

Marks, M. A., Mathieu, J. E., & Zaccaro, S. J. (2001). A temporally based framework and taxonomy of team processes. *Academy of Management Review, 26*, 356–376. doi:10.2307/259182

Mathieu, J. E., Tannenbaum, S. I., Donsbach, J. S., & Alliger, G. M. (2014). A review and integration of team composition models: Moving toward a dynamic and temporal framework. *Journal of Management, 40*, 130-160. doi:10.1177/0149206313503014

Moreland, R. L., & McMinn, J. G. (2010). Group reflexivity and performance. *Advances in Group Processes, 27*, 63–95. doi:10.1108/s0882-6145(2010)0000027006

Nederveen Pieterse, A., van Knippenberg, D., & van Ginkel, W. P. (2011). Diversity in goal orientation, team reflexivity, and team performance. *Organizational Behavior and Human Decision Processes, 114*, 153-164. doi:10.1016/j.obhdp.2010.11.003

Nickel, R. (2011). *Aristoteles: Die Nikomachische Ethik: Griechisch – Deutsch (Sammlung Tusculum)*. Düsseldorf: Artemis & Winkler.

Oertel, R., & Antoni, C. H. (2015). Phase-specific relationships between team learning processes and transactive memory development. *European Journal of Work and Organizational Psychology*, *24*, 726-741. doi:10.1080/1359432X.2014.1000872

Okhuysen, G. A. (2001). Structuring change: Familiarity and formal interventions in problem-solving groups. *Academy of Management Journal, 44*, 794-808. doi:10.2307/3069416

Okhuysen, G. A., & Waller, M. J. (2002). Focusing on midpoint transitions: An analysis of boundary conditions. *Academy of Management Journal, 45*, 1056-1065. doi:10.2307/3069330

Podsakoff, P. M., & Organ, D. W. (1986). Self-reports in organizational research: Problems and prospects. *Journal of Management, 12*, 531–544. doi:10.1177/014920638601200408

Romanelli, E., & Tushman, M. L. (1994). Organizational transformation as punctuated equilibrium: An empirical test. *Academy of Management Journal, 37*(5), 1141-1166.

Röd, W. (2000). *Der Weg der Philosophie Bd. 1: Altertum, Mittelalter, Renaissance*. München: C. H. Beck.

Savelsbergh, C. M., van der Heijden, B. I., & Poell, R. F. (2009). The development and empirical validation of a multidimensional measurement instrument for team learning behaviors. *Small Group Research, 40*, 578-607. doi:10.1177/1046496409340055

Schippers, M. C., Den Hartog, D. N., & Koopman, P. L. (2007). Reflexivity in teams: A measure and correlates. *Applied Psychology: An International Review, 56*, 189–211. doi:10.1111/apps.2007.56.issue-2

Schippers, M. C., Den Hartog, D. N., Koopman, P. L., & Wienk, J. A. (2003). Diversity and team outcomes: The moderating effects of outcome interdependence and group longevity and the mediating effect of reflexivity. Jo*urnal of Organizational Behavior, 24*, 779-802. doi:10.1002/job.220

Schippers, M. C., Den Hartog, D. N., Koopman, P. L., & van Knippenberg, D. (2008). The role of transformational leadership in enhancing team reflexivity. *Human Relations, 61*, 1593–1616. doi:10.1177/0018726708096639

Schippers, M. C., Edmondson, A. C., & West, M. A. (2014). Team reflexivity as an antidote to team information-processing failures. *Small Group Research, 56*, 731–769. doi:10.1177/1046496414553473

Schippers, M. C., Homan, A. C., & van Knippenberg, D. (2013). To reflect or not to reflect: Prior team performance as boundary condition of the effects of reflexivity on learning and final team performance. *Journal of Organizational Behavior, 34*, 6–23. doi:10.1002/job.1784

Schippers, M. C., West, M. A., & Dawson, J. F. (2015). Team reflexivity and innovation: The moderating role of team context. *Journal of Management, 41*, 769–788. doi:10.1177/0149206312441210

Schön, D.A. (1983). *The Reflective Practitioner: How Professionals Think in Action*. New York: Basic Books.

Shin, Y., Kim, M., & Lee, S. H. (2016). Reflection Toward Creativity: Team Reflexivity as a Linking Mechanism Between Team Goal Orientation and Team Creative Performance. *Journal of Business and Psychology*. Advance online publication. doi:10.1007/s10869-016-9462-9

Smith-Jentsch, K. A., Cannon-Bowers, J. A., Tannenbaum, S. I., & Salas, E. (2008). Guided team self-correction: Impacts on team mental models, processes, and effectiveness. *Small Group Research, 39*, 303-327. doi:10.1177/1046496408317794

Starbuck, W. (1965). Organizational growth and development. In J. G. March (Ed.), *Handbook of Organizations* (pp. 451-533). Chicago: Rand McNally.

Stasser, G., & Titus, W. (1985). Pooling of unshared information in group decision making: Biased information sampling during discussion. *Journal of Personality and Social Psychology, 48*, 1467-1478. doi:10.1037/0022-3514.48.6.1467

Tannenbaum, S. I., & Cerasoli, C. P. (2013). Do team and invidiaul debriefs enhance performance? A meta-analysis. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 55*, 231-245. doi:10.1177/0018720812448394

Tjosvold, D., Hui, C., & Yu, Z. (2003). Conflict management and task reflexivity for team in-role and extra-role performance in China. *International Journal of Conflict Management, 14*, 141-163. doi:10.1108/eb022895

Tjosvold, D., Tang, M. M., & West, M. (2004). Reflexivity for team innovation in China the contribution of goal interdependence. *Group & Organization Management, 29*, 540-559. doi:10.1177/1059601103254911

Tushman, M. L., & Romanelli, E. (1985). Organizational evolution: A metamorphosis model of convergence and reorientation. *Research in Organizational Behavior, 7*, 171-222.

Tushman, M. L., & Rosenkopf, L. (1992). Organizational determinants of technological-change-toward a sociology of technological evolution. *Research in Organizational Behavior, 14*, 311-347.

Van Dick, R., & West, M. A. (2005). *Teamwork, Teamdiagnose, Teamentwicklung: Praxis der Personalpsychologie. [Teamwork, Team diagnosis, Team development*. Practices in Personnel psychology]. Göttingen: Hogrefe.

van Ginkel, W. P., & van Knippenberg, D. (2009). Knowledge about the distribution of information and group decision making: When and why does it work? *Organizational Behavior and Human Decision Processes*, *108*, 218-229. doi:10.1016/j.obhdp.2008.10.003

van Ginkel, W.P., Tindale, R.S. & van Knippenberg, D. (2009). Team reflexivity, development of shared task representations, and the use of distributed Information in group decision making. *Group Dynamics: Theory, Research and Practice, 13*, 265-280. doi:10.1037/a0016045

West, M. A. (1996). Reflexivity and work group effectiveness: A conceptual integration. In M. A. West (Ed.), *Handbook of Work Group Psychology* (pp. 555–579). Chichester, UK: John Wiley & Sons.

West, M. A. (2000). Reflexivity, revolution, and innovation in work teams. In M. M. Beyerlein, D. Johnson, & S. T. Beyerlein (Eds.), *Product Development Teams* (Vol. 5, pp. 1–29). Stamford, CT: JAI Press.

Wiedow, A., & Konradt, U. (2011). Two-dimensional structure of team process improvement: Team reflection and team adaptation. *Small Group Research, 42*, 32–54. doi:10.1177/1046496410377358

Wiedow, A., Konradt, U., Ellwart, T., & Steenfatt, C. (2013). Direct and indirect effects of team learning on team outcomes: A multiple mediator analysis. *Group Dynamics: Theory, Research, and Practice, 17*, 232-251. doi:10.1037/a0034149

Zajac, S., Gregory, M. E., Bedwell, W. L., Kramer, W. S., & Salas, E. (2014). The cognitive underpinnings of adaptive team performance in ill-defined task situations: A closer look at team cognition. *Organizational Psychology Review, 4*, 49-73. doi:10.1177/2041386613492787

Kapitel 2: Studie 1

Development and Validation of the REMINT: A Reflection Measure for Individuals and Teams

Otte, K.-P., Konradt, U., Garbers, Y., & Schippers, M. C. (2017). Development and validation of the REMINT: a reflection measure for individuals and teams. *European Journal of Work and Organizational Psychology*, *26,* 299-313. doi:10.1080/1359432X.2016.1261826

Abstract

A growing number of studies have investigated the role of team reflexivity, the extent to which teams reflect on and adapt their functioning. However, the way team reflexivity has been conceptualized and operationalized reveals several weaknesses, in particular the conception as a unidimensional construct. To provide greater conceptual clarity, we therefore propose a team reflexivity framework that integrates four interacting but distinct reflexive processes. In four studies, we focus on reflection as a fundamental reflexive process, and develop and validate an extended multidimensional reflection measure that captures the relevant dimensions of quality and quantity of reflection and the key transition processes of information seeking and information evaluation. Moreover, in order to delineate two common composition methods, we develop and validate a direct consensus and a referent-shift consensus version of the reflection measure. Data collected from a total of 803 students and employees in four studies revealed excellent construct validity, as well as good nomological validity (Studies 1 and 2). Furthermore, we found evidence of the criterion-related validity at the team level (Study 3) and the individual level (Study 4). Together, the results demonstrate the effectiveness of our measure, revealing consistent relations with outcome measures and diverse behavioural indicators across different contexts.

2.1 Introduction

Research on team reflexivity, -- the extent to which teams reflect on objectives, strategies, and processes and implement changes as a result (West, 2000) --, has revealed that reflexivity can have beneficial effects on team outcomes, but that it can also have detrimental or neutral effects (for a review, see Konradt, Otte, Schippers, & Steenfatt, 2016). This led some scholars to question the relevance and usefulness of the concept (Moreland & McMinn, 2010). Providing an alternative perspective, some recent reviews of this literature (Konradt et al., 2016) have argued that contradictory outcomes may arise from measures that do not capture the construct of interest sufficiently well and that this can lead to construct underrepresentation (Messick, 1995).

As Konradt and colleagues have argued, reflexivity has been largely conceived of as global and unidimensional construct, ignoring its potential multidimensional nature. In particular, reflexivity measures are restricted to a mainly quantitative construct (i.e., amount and frequency), in which other aspects such as specificity and depth are not represented (see Moreland & McMinn, 2010). Moreover, the measures of reflexivity currently available do not integrate different behavioural types and forms of reflexivity, such as feedback seeking (Anseel, Lievens, & Schollaert, 2009) and evaluation (Schippers, Den Hartog, & Koopman, 2007; Schippers, Edmondson, & West, 2014). Consequently, they do not cover all the aspects of reflexivity required to make a comprehensive interpretation of the behaviours involved, resulting in a loss of content validity, incorrect predictions and inconsistent results. Ultimately, this also hinders the development of more elaborated reflexivity theory and hypotheses as well as valid practical interventions.

In this study, we therefore argue that because of the variety of reflexive behaviours, we need to develop more advanced ways of measuring them. Moreover, we conceive of reflexivity as a framework that integrates four distinct reflexive processes. By building upon

previous research that focussed on reflection as a core element of reflexivity and integrating frameworks of team effectiveness (Ilgen, Hollenbeck, Johnson, & Jundt, 2005; Marks, Mathieu, & Zaccaro, 2001), we develop and validate a richer and more fine-grained measure which considers reflection in terms of the quality and quantity of information seeking and information evaluation. To create a reliable and valid measure of reflection, which we have called the 'Reflection Measure for Individuals and Teams' (REMINT), we followed research strategies outlined by Churchill (1979), DeVellis (2003), and Hinkin (1998), and we present our results in four studies. In Study 1, we describe the development process and examine the construct validity of the REMINT using confirmatory factor analysis (CFA) with two large student samples. In Study 2, we use an employee sample to further examine the construct validity within a third CFA and we establish a nomological network, which includes multiple constructs similar to the REMINT. In Studies 3 and 4, we examine the criterion-related validity of the REMINT longitudinally and provide evidence for the validity of both team and individually framed versions.

2.2 Theoretical Background and Hypotheses

It is worth mentioning at the start that previous research sometimes did not distinguish between *team reflection* and *team reflexivity*, but used the terms interchangeably. Team reflection has been conceptualized in a variety of ways (for an extensive discussion, see Gabelica, Van den Bossche, Segers, & Gijselaers, 2014; Schippers, Edmondson, & West, forthcoming). In this research, we refer to the definition by Savelsbergh, van der Heijden, and Poell (2009), who defined reflection as looking back on experiences, goals, actions, working methods, strategies, and assumptions.

The team reflexivity construct is a much broader in concept than team reflection. Team reflexivity refers to the "extent to which team members collectively reflect upon the team's objectives, strategies, and processes, as well as their wider organizations and

environments, and adapt them accordingly" (West, 2000, p. 3). In his definition of team reflexivity, West emphasizes planning and adaptation as elements beside reflection that together constitute an optimal sequence of reflection, planning and action/adaptation.

Thus, most research has focused on the reflection element, with planning being largely neglected (see for an exception Gevers, van Eerde, & Rutte, 2001) and only a few articles that have included adaptation (Schippers et al., 2007; Wiedow & Konradt, 2011; Konradt, Schippers, Garbers, & Steenfatt, 2015). In this paper, we argue that reflection is the core reflexivity process. We will therefore focus on reflection and its components. We do not want to deny the importance of planning or adaptation/implementation, but rather argue that understanding the full significance of both constructs would require extensive additional work that is beyond the scope of this study.

When reviewing measures of reflection, Konradt et al. (2016) concluded that previous scales focussed mainly on the quantity or extent of reflection, as indicated by the *amount or number* of related behaviours. Qualitative aspects relating to how deep or how detailed reflexive behaviours may be (i.e., the depth of information processing, see Schippers et al., 2014) are not captured by current versions of team reflection questionnaires, and Schippers et al. (2014) conclude that "there would be considerable value in understanding how to conceptualize depth of reflexivity" (p. 754). Several authors also argue that assessing quality in team reflection constructs could be more relevant than assessing quantity (Moreland & McMinn, 2010; Schippers et al., 2014). Moreover, strength on one dimension might be achieved at the expense of the other. A qualitative focus might prevent teams from covering a wide range of different subjects, but might enable them to elaborate in more detail on selected topics. On the other hand, a quantitative focus might compromise quality and have a negative impact on the outcomes (e.g., performance), because teams may come to wrong or incomplete conclusions. As a case in point, a team might spend a great deal of time on reflection but the

reflection might remain shallow and unsubstantial, touching only superficially on points, without finding the underlying causes. Hence, the time invested might not result in better team decision-making and performance. We therefore argue that it is crucial to make a distinction between both quality and quantity in order to cover all relevant team behaviours and make accurate predictions of how reflection relates to different outcomes (c.f. Schippers et al., 2014).

In addition to the distinction between quantity and quality, prior research has also revealed that different types of behaviour are particularly important for understanding the relationship between reflection and performance in teams. This is important because several experimental studies have shown that reflection on its own does not lead to improvements in performance. Instead, feedback on the success of the methods used, for example, needs to be considered as another substantial element. Anseel et al. (2009), for example, showed that combining (individual) feedback with reflection leads to greater performance improvements on a work-related simulation task than either feedback or reflection. Gabelica, Van den Bossche, De Maeyer, Segers, and Gijselaers (2014) and Konradt et al. (2015) generalized this finding at the team level by showing that a combination of team performance feedback and guided reflection led to the greatest improvements in team performance. Field studies also suggest there are positive relationships between feedback and reflection (Schippers et al., 2007; Savelsbergh et al., 2009) and Schippers et al. (2014) argued that searching for information is an important aspect of team reflection.

Because in many instances feedback is simply not available or salient, we argue that seeking out information on how the team is functioning is crucial if the team is to benefit from reflection (Konradt et al., 2016). Hence, the degree to which teams seek out information needs to be integrated into the reflexivity framework as an additional element.

To conclude, we propose three main adjustments to the way that reflexivity is conceived and measured. First, drawing on West's (2000) work, we conceive of reflexivity as a framework (as opposed to a single construct as previously) that integrates four *reflexive processes*, namely (1) seeking information on the functioning of the team (i.e., the methods used, strategies, processes, etc.); (2) evaluating this information; (3) planning what action to take on the basis of these evaluations; and (4) implementing the planned actions (c.f. adaptation in West, 2000; Schippers et al., 2007). It is important to state that while there is interaction between these processes, each should be measured individually and should therefore also be treated as distinct process in the reflexivity framework and statistical analysis.

Second, we position these processes within the team transition/action phase of the framework developed by Marks et al. (2001) in order to provide a better understanding of when and why reflexive processes typically (should) occur. Transition phases are "periods of time when teams focus primarily on evaluation and/or planning activities to guide their accomplishment of a team goal or objective" (Marks et al., 2001, p. 360). Action phases are "periods of time when teams are engaged in acts that contribute directly to goal accomplishment (i.e., task work)" (Marks et al., 2001, p. 360). Third, it is worth noting that each of these four processes have both a quantitative and a qualitative dimension. Thus, as indicated earlier, in this research we will focus on the two key elements of information seeking and information evaluation in terms of both quantity and quality.

Information seeking refers to gathering information from various sources, returning unprocessed information on the status of the team (e.g., gathering information on successes and failures). Information seeking involves monitoring activities (i.e., receiving information through passive observations and inferences from the environment) and inquiring activities (i.e., asking directly for information or feedback from colleagues or superiors) (see the

feedback-seeking literature such as Anseel, Beatty, Shen, Lievens, & Sackett, 2015). It can be categorized as a process, which takes place during action phases (see also various monitoring activities in Marks et al., 2001), but can also occur during transition phases when additional information is required.

The second aspect, information evaluation, is considered to be a central part of reflexivity, also often referred to as reflection (Schippers et al., 2014) or evaluation/learning (Schippers et al., 2007). Information evaluation means aggregating pieces of information and making interpretations that go beyond their individual significance in order to draw more elaborate conclusions and develop new or different methods and strategies. We argue that the qualitative sub-dimension of information evaluation is particularly important, because deeper processing requires more time and effort and is not possible when the discussion remains at a superficial level (i.e., quantitative focus). We categorize information evaluation as a transition process which takes place between consecutive action phases.

Hypothesis 1: The newly designed measure of reflection is best conceptualized as having four dimensions: Information-seeking (quantity and quality) and information evaluation (quantity and quality).

This more fine-grained conception of reflection enables us to look in more detail at what processes take place within teams. In particular, teams that have a purely quantitative (and therefore possibly harmful) focus on evaluation, for example, could be identified, and effective interventions developed to address this. By integrating different streams of research (feedback-seeking, reflection) our approach supports the development of richer and more detailed theories. This becomes especially relevant because changes in reflection over time (Konradt & Eckardt, 2016) and the level of measurement (i.e., individual and team) are crucial for understanding team dynamics (Humphrey & Aime, 2014).

In addition, group-level constructs that are derived from individual-level data are based on a composition model that specifies how the lower level data can be combined to form the higher-level construct (Rousseau, 1985). Thus, based on Chan's (1998) proposed typology of composition models, we develop two versions of our measure, which reflect both direct consensus composition, where items refer to the individual, and referent-shift consensus composition, where items refer to the group.

2.3 Nomological Network

In order to validate the claim that a test measures a certain construct, "a nomological net surrounding the concept must exist" (Cronbach & Meehl, 1955, p. 291). A nomological network consists of the primary construct being investigated and several other constructs that are positively (convergent validity) related to that primary construct or unrelated to it (divergent validity). Moreover, other criteria may be involved that are predicted by the primary construct (e.g., outcomes or consequences). In the following, we will therefore discuss several measures related to team reflection.

Because the REMINT is a reflection measure, there will be overlap with previous measures of team reflection/reflexivity. Though there are already many scales that measure reflection and reflexivity (for an overview of existing scales, see Konradt et al., 2016), most originate from Carter and West's (1998) Task Reflexivity Scale, which was later relabelled reflexivity (West, 2000). Schippers et al. (2007) enriched West's measure by incorporating a wider range of items to measure reflection that measure the two dimensions evaluation/learning and discussing processes. In the current paper we use these two measures (West, 2000, and Schippers et al., 2007) to examine convergent validity. Since both measures focus mainly on quantitative measurements of reflection, we predict that the two scales will correlate more strongly with the quantitative dimensions of the REMINT than with the qualitative dimensions. In addition, as the Schippers et al. (2007) scale represented a

substantial refinement in terms of measuring reflection, correlations between the REMINT and that scale should be higher than correlations between the REMINT and the West (2000) scale.

Hypothesis 2a: The REMINT scales are positively related to West's (2000) and Schippers et al. (2007) reflexivity scales.

Hypothesis 2b: The quantitative subscales of the REMINT are more positively related to the West (2000) and Schippers et al. (2007) scale than the qualitative subscales.

Hypothesis 2c: The REMINT scales are more positively related to the Schippers et al. (2007) scale than to the West (2000) scale.

To further examine convergent validity in the team context, we use psychological constructs in the team context that are similar to reflection, including Edmondson's (1999) team learning concept, information sharing (De Dreu, 2007), feedback-seeking behaviours (Anseel et al., 2015), and team monitoring (De Jong & Elfring, 2010). Team learning refers to processes of group interaction activities through which individuals acquire, share, and combine knowledge (Edmondson, 1999). The construct shares many similar characteristics with reflection (for an elaboration, see Schippers et al., 2013) and is expected to show a positive relationship with all dimensions of the REMINT. Because feedback-seeking behaviours (FSB) – defined as "individual actions to gather information relevant to one's own behaviour" (Anseel et al., 2015, p. 319) – are similar to elements in our measure, we expect there to be positive correlations between the information-seeking subscales of the REMINT and FSB. Information sharing (De Dreu, 2007) refers to the level or amount of information shared within a team. Hence, reflecting within a team should be based to a large extent on discussing things (ie, sharing information). In a cross-sectional field study with 368 individuals in 46 teams, De Dreu reported a correlation of $r = .48$ to West's measure. We therefore expect the information evaluation subscales of the REMINT are positively related to

the information-sharing measure. Finally, team monitoring (De Jong & Elfring, 2010) involves checking whether team members are performing as expected. Monitoring teamwork also means gathering information on particular aspects of the team's functioning. We therefore expect there to be a little relationship between team monitoring behaviours and the information-seeking dimensions of the REMINT. Thus, this will not be the case for the information evaluation dimensions, where we anticipate no correlation with team monitoring (De Jong & Elfring reported a correlation of .63 to West's reflexivity measure).

Hypothesis 3a: The team version of the REMINT is positively related to team learning behaviours.

Hypothesis 3b: The information-seeking subscales of the team version of the REMINT are positively related to team FSB, while the information evaluating subscales are not.

Hypothesis 3c: The information evaluation subscales of the team version of the REMINT are positively related to team-level information sharing, while the information-seeking subscales are not.

Hypothesis 3d: The information-seeking dimensions of the team version of the REMINT are positively related to team monitoring. The information evaluation dimensions of the team version of the REMINT are not correlated to team monitoring.

West's reflexivity concept assumes that reflexive processes are positively related to outcomes such as performance, effectiveness, or innovation (see Schippers, West, & Edmondson, in press). Thus, any reflection measure needs to show whether and how it predicts outcomes and how it is indicative of predictive validity (Nunnally & Bernstein, 1994). We thus expect that the REMINT will predict performance both at the individual and the team level.

Hypothesis 4: At the individual level, the REMINT is positively related to individual performance, and at the team level it is positively related to team performance.

2.4 Study I: Item Generation and Construct Validity

Following established scale development guidelines (Churchill, 1979; DeVellis, 2003; Hinkin, 1998), we used a systematic multiple-step process, including conceptual development, refinement, and validation. The items used on the scale need not only to meet the appropriate psychometric standards, but also to be suitable for use in a variety of field and experimental contexts (e.g., academic and non-academic), while still being as small a set as possible. As a first step, we created an item pool, using an analytical and a synthesizing approach. In the analytical approach, a group of six experts in work and organizational psychology independently reviewed the existing measures of reflexivity/reflection/feedback-seeking behaviour and identified items that they found suitable for measuring information seeking and information evaluation. Next, in the synthesizing approach, the same group of experts created set of items based on two lists of prototypical verbs (e.g., evaluation, analysing, searching, or discussing) and substantives (e.g., strategies, processes, problems, or successes) related to information-seeking and information evaluation activities. These experts then rated the resulting items in terms of whether they adequately represented these particular dimensions and the associated behaviours. Any items felt not to be appropriate were removed and new items were added. Finally, we grouped the items by dimension and reduced the number of items to eight for each dimension, selecting among those with similar wording to avoid duplication. We created quantitative and qualitative subsets of items by adding terms either indicating quantity (e.g., “very often”) or quality (e.g., “in great detail”), allowing the use of a response format ranging from “strongly disagree” to “strongly agree”. The resulting set of 32 items was used to assess the construct validity of the REMINT within samples of the target populations (see also Study 2).

2.4.1 Method

2.4.1.1 Participants and Procedure.

Cognitive interviews (Fisher & Geiselman, 1992) were conducted with eight undergraduate students and eight employees in order to verify whether the items and the response format were clear and whether they could be adjusted to suit to individual work settings. Items were adapted slightly based on their comments. Next, we recruited 400 undergraduate students (mean age = 23.25, *SD* = 3.87, 68.3 % female) at two German universities. Hoelter (1983) and Hinkin (1998) suggest that confirmatory factor analysis (CFA) require at least 200 participants (see also MacCallum, Widaman, Zhang, & Hong, 1999). Following Hinkin's (1998) suggestion, we split the sample randomly in two halves and conducted parallel analysis on both datasets ($n_1 = n_2 = 200$).

Participants were sent an invitation email with a link to an online survey. They were told to recall a certain task (for example, working in a project team for a university course) that they were either currently working on or had worked on within a team during the previous three months, and were instructed to answer all items relating to this task. The survey took about 20 minutes to complete, because other measures were also included. Afterwards, participants were thanked for their efforts and received either credit in the form of research participation points when studying psychology, or the chance to win one of four €25 vouchers for an online retailer.

2.4.1.2 Measure.

A set of 32 items framed at the team level was employed in this study. The response format consisted of a five-point Likert scale ranging from *strongly disagree* (1) to *strongly agree* (5). To avoid method effects and artificial latent factors in factor analysis (cf. Brown, 2015), all items were positively worded.

2.4.1.3 Data Analysis.

We tested for violations against normality assumptions (using the Shapiro–Wilk test), to avoid potential misjudgements and underestimations of fit statistics, if not accounted for (Satorra, 2000). We also checked whether skewness and kurtosis were at values greater or smaller than .5 (Bulmer, 1979). We tested the factorial validity with confirmatory factor analysis (CFA) using MPlus 7.3 (Muthén & Muthén, 2015). We analysed the variance–covariance matrix with a robust maximum likelihood estimator (MLR) (Satorra & Bentler, 1994) since normal distribution of the responses was not a given (as indicated by Shapiro–Wilk tests). We judged the model by evaluating the χ^2 goodness of fit statistic, the Comparative Fit Index (CFI), the Tucker–Lewis index (TLI), the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR). We followed relevant guidelines for goodness of fit (e.g., West, Taylor, & Wu, 2012), which state that a CFI and TLI of more than .95 and an RMSEA and SRMR of less than .05 represent an acceptable model fit. We used the scaled difference chi-square test statistic (Bryant & Satorra, 2012) to compare the χ^2 goodness of fit values of nested models, since the common way of comparing these values does not yield correct results when the MLR estimator is used (Satorra, 2000).

We started the factorial analysis by examining separate one-factor models for each of the four dimensions; this was followed by two-factor models separated out the information seeking and information evaluating dimensions, and finally by a four-factor solution which represented all four dimensions. Error terms of items measuring the same behaviour were allowed to covary. Furthermore, we evaluated the fit of the items within their respective dimensions by assessing factor loadings (items were deleted when $\lambda < .50$), standard errors, and modification indices (items were deleted when the modification indices indicated there were too many or too high error term correlations with other items that could not be explained

by the theoretical assumptions proposed prior to the statistical analysis; see MacCallum, 1986). If an item was removed from a quantitative subscale, we also removed it from the qualitative scale (or vice versa). We maximized content validity by keeping as much variance as possible regarding the content of the items.

Finally, when interpreting standardized factor loadings, we followed Comrey and Lee's (1992) recommendations ($\lambda > .71$ = excellent; $\lambda > .63$ = very good; $\lambda > .55$ = good; $\lambda > .45$ = fair; $\lambda < .32$ = poor). We calculated the reliability coefficient alpha to estimate internal consistency. Discriminant validity was examined by analysing factor correlations, which should not exceed values greater than .85 (cf. Cohen, Cohen, West, & Aiken, 2003).

2.4.2 Results and Discussion

The analysis reported below concerns the calibration sample ($n_1 = 200$). Following the procedures described above, our first step was to reduce the initial item set from 32 items to 20 items, with five items representing each subscale. Each subscale showed excellent fit values, exceeding the requirements for good fit. We then combined the reduced subscales into our final four-factor model (quantity and quality of information seeking and information evaluation). Analysis showed that correlating the error terms of the qualitative and quantitative items was justified and led to great improvements in the model fit values. In the next step, we reduced the item set from 20 to 16 items, removing all items with factorial loadings below .50. The final item set is displayed in Table 2.1.

Our proposed four-factor model yielded significantly better fit values (see Table 2.2) than a unidimensional model (with all items loading on a single factor) and a two-dimensional model (that did not include the quality and quantity differentiation). We also tested a higher-order reflection factor which combined all four dimensions into one factor, but this solution yielded significantly worse fit values than the originally proposed four-dimensional first-order solution. Furthermore, combining both information seeking and information evaluation

factors into two higher-order factors resulted in a non-positive definite latent factor matrix with several latent factor correlations greater than 1 (positive-definite matrices are a requirement for structural equation modelling).

Factor loadings (λ) of our proposed four-factor model ranged from good to excellent, with the lowest λ being .64 and the highest λ being .83 with an excellent average loading of .73. Reliability values were .78 and .83 for both information-seeking factors and .82 and .86 for information evaluation factors. Model fit indices were very good, with both goodness-of-fit indices CFI and TLI above the set threshold of .95 (both .97) and badness-of-fit indices SRMR (.04) and RMSEA (.04) below the threshold of .05. The model reached a χ^2 (90) of 122.07, $p < .05$.

Discriminant validity was supported, with the exception of the correlation between quantity and quality factors in the information-seeking dimension, which was slightly higher than .85 (cf. Cohen et al., 2003). For the validation sample, the results from the calibration sample were replicated, with mainly equivalent fit indices and factor loadings, as displayed in Table 2.1 and Table 2.2.

Overall, both samples supported the construct validity of the REMINT and thereby our proposed model structure for measuring two of the four elements of reflexivity. Consistent with our first hypothesis (H1), a four-factor model comprising quantity and quality of information-seeking and information evaluation factors yielded the best fit when compared to other models. The fact that a four-factor solution yielded a better fit than a second-order general factor solution suggests that our measure shows construct validity. One critical aspect pertains to the (partly missing) discriminant validity in the validation sample, since at least the latent factor correlations between the quantitative and qualitative information-seeking factor are higher than the previously set value of .85. Other latent factor correlations also reached high values, with multiple values exceeding .85 within the validation sample. This difficulty

Table 1.1: Standardized Item Loadings of the REMINT with Standard Errors in Parenthesis

Factor	No	Item	Validation Sample (Study 1; *n* = 200)	Calibration Sample (Study 1; *n* = 200)	Employee Sample (Study 2; n = 93)	Individual Sample (Study 4; *n* = 163)
IS Quantity	1	We often tried to find the cause of problems in our work.	.64 (.05)	.56 (.07)	.46 (.10)	.62 (.07)
	2	We often observed the consequences of our work.	.77 (.04)	.75 (.04)	.63 (.09)	.75 (.05)
	3	We collected information about our successes and failures very frequently.	.66 (.05)	.68 (.06)	.62 (.09)	.66 (.05)
	4	We monitored the progress of our work very frequently.	.67 (.05)	.70 (.05)	.61 (.09)	.63 (.07)
IS Quality	5	We went into great detail to find the cause of problems in our work.	.70 (.05)	.77 (.04)	.68 (.07)	.70 (.05)
	6	We observed the consequences of our work in great detail.	.74 (.05)	.75 (.04)	.66 (.07)	.78 (.04)
	7	We collected very detailed information about our successes and failures.	.74 (.04)	.74 (.05)	.74 (.06)	.67 (.05)
	8	We monitored the progress of our work in great detail.	.78 (.04)	.82 (.03)	.65 (.07)	.80 (.04)
IE Quantity	9	We made very frequent assessments of how successful our procedures were.	.73 (.05)	.71 (.05)	.67 (.08)	.80 (.05)
	10	We made very frequent evaluations of the quality of our work.	.69 (.05)	.69 (.05)	.47 (.10)	.77 (.05)
	11	While engaged in tasks, we checked very frequently whether we were on the right track.	.70 (.05)	.66 (.06)	.55 (.10)	.66 (.06)
	12	We evaluated our progress during our work very frequently.	.79 (.03)	.70 (.05)	.70 (.08)	.67 (.06)
IE Quality	13	We made very detailed assessments of how successful our procedures were	.82 (.03)	.76 (.04)	.85 (.05)	.83 (.03)
	14	We made very detailed evaluations of the quality of our work.	.76 (.04)	.72 (.04)	.78 (.05)	.76 (.05)
	15	While engaged in tasks, we checked in great detail whether we were on the right track.	.71 (.04)	.70 (.05)	.71 (.06)	.77 (.04)
	16	We evaluated our progress during our work in great detail.	.83 (.03)	.81 (.03)	.70 (.07)	.71 (.06)

Note. IS = Information seeking. IE = Information evaluation. Estimators = MLR (robust maximum likelihood) and Bayes (employee sample). All loadings were significant at $p < .001$.

Table 2.2: Fit Indices of the CFA, Using Calibration Data, Validation Data, Employee Data, and Individual-level Data

	χ^2	*df*	*p* <	CFI	TLI	SRMR	RMSEA [90% CI]	χ^2 *Diff*	PPP	PPC CI	BIC
Model 1: 1 Factor (Reflection)											
Validation sample	219.07	96	.001	.919	.899	.057	.080 [.066; .094]	––	––	––	8119.23
Calibration sample	197.26	96	.001	.926	.908	.055	.073 [.058; .087]	––	––	––	8209.66
Employee sample	––	56	––	––	––	––	––	––	.00	12.07; 402.89	4176.26
Individual Sample	200.35	96	.001	.908	.885	.055	.082 [.066; .098]	––	––	––	6268.71
Model 2: 2 Factors (Information seeking/evaluating)											
Validation sample	174.48	95	.001	.948	.934	.047	.065 [.049; .079]	$p < .001$	––	––	8072.00
Calibration sample	157.78	95	.001	.954	.942	.046	.058 [.041; .073]	$p < .001$	––	––	8168.34
Employee sample	––	57	––	––	––	––	––	––	.00	32.97; 1409.79	3810.96
Individual sample	171.65	95	.001	.933	.915	.047	.070 [.053; .087]	$p < .001$	––	––	6238.93
Model 3: 1 + 4 Factors (Reflection + Information seeking/evaluating quantity/quality)											
Validation sample	137.69	92	.001	.970	.961	.046	.050 [.031; .066]	$p < .001$	––	––	8044.14
Calibration sample	135.90	92	.002	.968	.958	.048	.049 [.030; .066]	$p = .003$	––	––	8154.29
Employee sample	––	60	––	––	––	––	––	––	.12	-20.32; 177.28	3795.00
Individual sample	119.81	92	.027	.976	.968	.043	.043 [.015; .063]	$p < .001$	––	––	6190.05
Model 4: 4 Factors (Information seeking/evaluating quantity/quality)											
Validation sample	122.07	90	.014	.979	.972	.041	.042 [.020; .060]	$p < .001$	––	––	8037.19
Calibration sample	118.86	90	.022	.979	.972	.043	.040 [.016; .059]	$p < .001$	––	––	8145.11
Employee sample	––	62	––	––	––	––	––	––	.39	-39.11; 49.25	3753.43
Individual sample	106.78	90	.11	.985	.980	.039	.034 [.000; .056]	$p < .001$	––	––	6185.28

Note. N is between 200 and 162. PPP = Posterior predictive p-value. PPC CI = Posterior predictive checking confidence interval. BIC = Bayesian information. χ^2 Diff = Satorra-Bentler scaled chi-square value for chi-square difference testing to compare a SEM model with a less restricted one.

might arise because of the samples that were used for the CFAs. Even though many students are experienced team workers, most student teams only collaborate for a short time, resulting in limited opportunities for deeper reflection and therefore perhaps only limited capabilities to differentiate between quantity and quality. Study 2 therefore used an employee sample to investigate this matter further.

2.5 Study 2: Replication and Nomological Network of the REMINT

In Study 2, we investigated whether the results from Study 1 would be generalizable to a different context of employees working in teams. Additionally, we tested whether reflection (as measured by the REMINT) related to convergent constructs/measures (i.e., previous reflexivity/reflection measures, team learning, team monitoring, feedback-seeking behaviours, and information sharing), for examining aspects of the nomological network around reflection. Hence, we recruited a new sample that included employees from various areas, and followed a similar procedure to that outlined for the previous study. We expected to replicate the fit between our model and the data found in Study 1, and the relations to all the constructs described previously in the theoretical background. A replication of the previous results and correlations within the expected ranges would further validate our claim that the REMINT measures the intended construct.

2.5.1 Method

2.5.1.1 Participants and Procedure.

The procedures were the same as reported in Study 1. We recruited 93 employees (mean age = 33.99, *SD* = 10.31, 53.26% female) from various areas (e.g., consulting, management, public service and research) currently working in teams (mean size = 3.93, *SD* = 1.73). All employees were German. Employees were recruited by means of personal contacts in various companies, mailing lists in these companies, social networks (LinkedIn and Xing) and research blogs. Twenty-one of these employees filled out only a short reflection

questionnaire, whereas the other 72 completed the same questionnaire as in Study 1.

2.5.1.2 Measures.

The response format for all measures was a five-point Likert scale ranging from *strongly disagree* (1) to *strongly agree* (5).

Reflection. Reflection was assessed using three scales: The REMINT, and the scales of West (2000) and Schippers et al. (2007). The qualitative sub-dimensions of the REMINT (information seeking and information evaluation) showed a better internal consistency (α = .77, and .84, respectively) than the quantitative dimensions (α = .67, and.70). We used van Dick and West's (2005) German translation of West's (2000) team reflexivity scale, with eight items (e.g., "We often reviewed the methods we used to get the job done"). Internal consistency was good (α = .84). We also used a shortened eleven-item version of the Schippers et al. (2007) scale (e.g., "We talk about different ways in which we can reach our objectives"), which had been translated into German by Brahm (2009). Internal reliability was acceptable (α = .79).

Team Learning. Team learning was assessed using the validated German version (Wiedow & Konradt, 2011) of Edmondson's original team learning measure (1999). Items included, for example, "We regularly take time to figure out ways to improve our team's work processes". The measure showed an acceptable internal consistency (α = .74).

Feedback-seeking Behaviours. Feedback-seeking behaviour was assessed using a five-item scale from Schippers et al. (2007) (e.g., "We seek feedback on our methods"). We chose this scale because it had already been deployed in previous scale validation of the Schippers et al. (2007) reflexivity measure. The scale was translated into German by a group of three experts who translated the items individually and then reached consensus in a follow-up discussion. The measure showed an acceptable internal consistency (α = .77).

Team Monitoring. Team monitoring (e.g., "In this team we check whether everyone

meets their obligations to the team") was measured using De Jong and Elfring's (2010) team monitoring scale. We chose this scale since it had already been used in a team reflexivity context. The translation into German was undertaken by a group of experts who translated the five items individually before reaching consensus in a follow-up discussion (α = .91).

Information Sharing. Information sharing was measured using De Dreu's (2007) five-item information-sharing scale (e.g., "Members of my team inform each other about work-related issues"). Like the other translated scales, the translation into German was carried out by a group of experts who translated the five items individually before reaching consensus in a follow-up discussion (α =.83).

2.5.1.3 Data analysis.

As the sample size of the employee sample (N = 93) was not large enough for common maximum likelihood estimations, we decided to use Bayesian structural equation modelling (BSEM) (see Muthén & Asparouhov, 2012; Zyphur & Oswald, 2015) to analyse the employee dataset. BSEM has several advantages over common maximum likelihood estimations, including the fact that it allows for a better handling of smaller sample sizes (Song & Lee, 2012). We employed two Markov chain Monte Carlo chains (Gibbs algorithm) and set the number of iterations to 100,000. To assess model fit, we used the posterior predictive p-value (PPP) and a posterior predictive checking (PPC) 95% credibility interval (CI). A PPP value above .05 indicates an acceptable fit, whereas a PPP value of .5 indicates an excellent fit (Muthén & Asparouhov, 2012). The PPC CI should include a negative lower limit and a positive upper limit (Zyphur & Oswald, 2015) to indicate a good fit. We also verified whether the solution was stable when we changed the number of iterations. The analysis was conducted in Mplus 7.4 (Muthén & Muthén, 2016).

2.5.2 Results and Discussion

The four-factor model provided a good fit to the data, with a PPP of .394 and PPC

95% CI of [-39.112 | 49.151] and these did not change when the number of iterations was altered. Factor loadings (see Table 2.1) ranged from .46 to .85, with a very good average loading of .66. Again, the four-factor model fitted the data better than the two- and one-factor models, as indicated by the lower BIC (see Table 2.2), further supporting the construct validity of the REMINT. Intercorrelations between the REMINT scales ranged from .39 ($p <$ .01) to .69 ($p < .001$), suggesting discriminant validity between the REMINT subscales. Reliability alpha (which ranged between .67 and .84) and correlations between the variables in the proposed nomological network, are presented in Table 2.3 ($n = 72$). Hypothesis 2a predicted a positive relationship between the REMINT and previous reflection/reflexivity scales. The hypothesis was confirmed, with correlations ranging between .26 ($p < .05$) and .53 ($p < .01$). Hypothesis 2b predicted a higher correlation between the quantitative scales of the REMINT and West's (2000) reflexivity measure than the qualitative scales of the REMINT and West's reflexivity measure. The hypothesis was confirmed, with both quantitative subscales of the REMINT and West's scale being more highly correlated than the qualitative subscale (see Table 2.3).

All correlations ranged between .26 ($p < .05$) and .40 ($p < .01$). Hypothesis 2c predicted a stronger relationship between the REMINT and the Schippers et al. (2007) reflexivity measure than between the REMINT and West's measure. This hypothesis was partly confirmed. All four subscales of the REMINT showed higher correlations with the Schippers et al. (2007) scale (ranging between .42 to .53; $p < .01$) than with the West (2000) scale (.26 to .40; $p < .05$). Testing for differences between the REMINT–West and the REMINT–Schippers correlations (Lee & Preacher, 2013) revealed significant (one-tailed according to the hypothesis) differences for the qualitative dimensions (information seeking quality $p < .05$; information evaluation quality $p < .01$), while the differences between the quantitative dimensions did not reach significant levels (information seeking quantity $p = .09$;

Table 2.3: Correlations and descriptive statistics for study variables in the employee sub-sample

	M (SD)	1	2	3	4	5	6	7	8	9	10
1 Information seeking quantity	3.51 (.71)	**.67**									
2 Information seeking quality	3.28 (.77)	.65**	**.77**								
3 Information evaluation quantity	3.46 (.60)	.39**	.53**	**.70**							
4 Information evaluation quality	3.27 (.77)	.47**	.69**	.56**	**.84**						
5 Reflexivity (sensu West)	3.10 (.66)	.29*	.26*	.40**	.31*	**.75**					
6 Reflexivity (sensu Schippers)	3.44 (.57)	.42**	.42**	.53**	.53**	.59**	**.79**				
7 Feedback-seeking behaviour	3.01 (.86)	.18	.19	.37**	.44**	.30*	.50**	**.77**			
8 Team learning	3.07 (.78)	.36**	.47**	.41**	.57**	.57**	.61**	.45**	**.74**		
9 Team monitoring	3.33 (.95)	.21	.11	.13	.00	.24*	.18	.18	.16	**.91**	
10 Information sharing	3.79 (.72)	.22[a]	.22[a]	.41**	.39**	.12	.50**	.50**	.44**	.14	**.83**

Note. $N = 72$. Reliability alphas on the diagonal (in bold).

[a]$p = .06$. *$p < .05$. **$p < .01$ (two-tailed).

information evaluation quantity $p = .07$). Hence, correlations between the qualitative dimensions of the REMINT and either West or Schippers differ significantly, while the quantitative dimensions do not.

Hypothesis 3a predicted a positive relationship between the REMINT and team learning behaviours. This was confirmed, with correlations between the REMINT and Edmondson's (1999) team learning scale ranging from .36 ($p < .01$) to .57 ($p < .001$). Hypothesis 3b, which predicted a positive relationship between the information-seeking subscales of the REMINT and FSB, and no correlation between the information evaluation subscales of the REMINT and FSB, was rejected. Correlations between the information-seeking subscales and FSB were not significant, while correlations between information evaluation subscales and FSB were .37 (quantity, $p < .01$) and .41 (quality, $p < .001$), respectively. Hypothesis 3c predicted a positive correlation between the information evaluation subscales of the REMINT and information sharing, and no correlation between the information-seeking subscales of the REMINT and information sharing. Our data supported this hypothesis. The information-seeking dimensions and information sharing showed a correlation of .22 (quantity and quality, $p = .06$) while the information evaluation dimension and information sharing showed a correlation of .41 (quantity, $p < .01$) and .39 (quality, $p < .01$). Hypothesis 3d, which predicted a positive relationship between the information-seeking subscales of the REMINT and team monitoring, and no correlation between information evaluation and team monitoring, received only partial support. Neither the information evaluation nor information-seeking subscales showed significant correlations with team monitoring.

Overall, four of our seven hypotheses relating to the nomological network were completely confirmed, two were partly confirmed, and one was not confirmed. These results provide further evidence of the construct and nomological validity of the REMINT, especially

because of the correlations with existing reflexivity scales. The Bayesian CFA further verified the dimensional structure of the REMINT found in Study 1. Additionally, the lower REMINT subscale intercorrelations suggest a much better discriminant validity than reported in Study 1, particularly between the qualitative and quantitative subscales. This result is very encouraging, because it indicates that our manipulation of quantity and quality was successful and that participants were able to distinguish between the two dimensions.

As proposed by Schippers et al. (2014) and Konradt et al. (2016), in Study 3 we used a longitudinal design to analyse the development of reflexivity and provide evidence of a relationship between the reflection measured by the REMINT and team (Study 3) and individual performance (Study 4).

2.6 Study 3: Team Reflection and Criterion-related Validity

The aim of Study 3 was to examine the relationship between reflection and performance at the team level. We used a two-wave longitudinal design to make a repeated assessment of team reflection and team performance.

2.6.1 Method

2.6.1.1 Participants and Procedures.

One-hundred forty-five graduate business management students (mean age = 25.3, *SD* = 3.06, 34% female) at a German university of applied science were randomly assigned to 34 teams (team size of between three and five members). The students were participating in a business simulation that ran for eight weeks, with each week representing one business year. Their task was to manage a fictional company and make complex decisions on relevant business operations in order to maximize the team's profit, while competing with other teams in the same dynamic market. During each business year, teams had to make several decisions and they later received feedback on the success of those decisions. The REMINT was deployed at business years 5 (t_1) and business year 6 (t_2), both times between making the

decisions and receiving feedback on the performance. These measurement points were chosen because an economic crisis occurred after business year 4, requiring teams to revise their strategies and adapt to new circumstances, a situation which real-life teams will often have to confront.

2.6.1.2 Measures.

The response format for all measures was a five-point Likert scale ranging from *strongly disagree* (1) to *strongly agree* (5).

Reflection. Reflection was measured using the reduced sixteen-item REMINT. Calculated *r*wg (j)* values (Lindell, Brandt, & Whitney, 1999) were above the required threshold (Smith-Crowe, Burke, Cohen, & Doveh, 2014), and thus responses were aggregated to the team level.

Team Performance. Team performance was measured as each company's surplus after each business year. The value was generated by the simulation software on the basis of several parameters (e.g., units sold, marketing budget, price of the product, or number of employees) which the teams had to plan for and make adjustments to during each business year.

2.6.2 Results and Discussion

Means, standard deviations, reliability alpha (which ranged between .64 and .91), $r^*wg_{(J)}$ values, and correlations are depicted in Table 2.4. Given the small sample, we conducted power analysis ($1-\beta$) with G*Power (Faul, Erdfelder, Buchner, & Lang, 2009) and used a level of significance of 10% to increase the power. All p-values are one-tailed, in line with Hypothesis 4. The results mainly support Hypothesis 4, which stated that information seeking and information evaluation would be positively related to team performance. At t_1, both the quantity ($r = .32$, $p = .03$, $R^2 = .10$, $1-\beta = .74$) and quality ($r = .34$, $p = .02$, $R^2 = .12$, $1-\beta = .79$) of team information evaluation predicted team performance, and also the quality

(but not the quantity) of team information seeking (r = .27, p = .06, R^2 = .07, $1-\beta$ = .63).

The same result was found in t_2 with the quality (but not the quantity) of team information seeking being significantly correlated to team performance (r = .29, p = .05, R^2 = .08, $1-\beta$ = .67). In contrast, only the quantity of team information evaluation was significantly correlated with team performance at t_2 (r = .37, p = .02, R^2 = .14, $1-\beta$ = .84). Overall, the results provide evidence of the criterion-related validity of the REMINT. The different results for t_1 and t_2 are surprising but clear. Since t_1 was the first measurement point after a significant change (economic crisis) which required teams to develop new strategies, deep and elaborate reflection was necessary. Afterwards at t_2, teams were mainly checking whether they had performed as expected, and were therefore gathering more information on their performance. This did not require deep, elaborate evaluation, which perhaps explains the results. Overall, the results should be regarded with caution because of the small sample size. Thus, this first team-level study supports Hypothesis 4 that the REMINT predicts team performance and provides evidence of the REMINT's criterion-related validity.

Another plausible and widely used composition model is the direct consensus model (Chan, 1998), in which responses are derived from individual-level data, but the referent is the individual ("I"). A direct consensus model uses "within-group consensus of the lower-level level units as the functional relationship to specify how a construct conceptualized and operationalized at the lower level is functionally isomorphic to another form of the construct at the higher level" (Chan, 1998, p. 237). Consequently, diversity in team-level constructs and individual-level influences can be delineated. The distinction between direct consensus and referent-shift consensus approaches in group-level constructs has rarely been considered to date (see van Mierlo, Vermunt, & Rutte, 2009, for a methodological framework), but we argue that, conceptually, both approaches can be regarded as appropriate for measuring reflection. We thus acknowledge the need for a direct consensus measure of reflection.

Table 2.4: Team-level Means, Standard Deviations (SD), r*wg(j) Agreement Indices, and Intercorrelations of the Team Sample

Variable	*M (SD)*	*r*wg(j)*	1	2	3	4	5	6	7	8	9	10	11
1 Team size	3.93 (.93)	na	na										
2 Performance t_1	–.17 (12.45)	na	–.05	na									
3 Performance t_2	–4.74 (15.04)	na	.19	.51**	na								
REMINT Time 1													
4 IS Quantity	3.88 (.37)	.68	.21	.13	.11	**.64**							
5 IS Quality	3.78 (.39)	.62	.13	.27*	.15	.66**	**.78**						
6 IE Quantity	3.76 (.41)	.65	.33**	.32**	.39**	.77**	.69**	**.86**					
7 IE Quality	3.63 (.50)	.63	.06	.34**	.25*	.61**	.88**	.80**	**.91**				
REMINT Time 2													
8 IS Quantity	3.88 (.37)	.68	.32**	.23*	.24*	.72**	.55**	.74**	.49**	**.85**			
9 IS Quality	3.79 (.47)	.62	.31**	.27*	.29**	.72**	.76**	.74**	.66**	.77**	**.92**		
10 IE Quantity	3.71 (.41)	.65	.24*	.23*	.37**	.74**	.54**	.77**	.57**	.78**	.76**	**.88**	
11 IE Quality	3.70 (.40)	.63	.23*	.32**	.20	.59**	.66**	.71**	.69**	.59**	.79**	.76**	**.89**

Note. N_{Team} = 35. REMINT = Reflection Measure for Individuals and Teams. IS = Information seeking. IE = Information evaluation. Reliability alphas on the diagonal (in bold).

*p < .10. **p < .05. All p-values are one-tailed according to the hypothesis.

2.7 Study 4: Individual Reflection and Criterion-related Validity

Study 4 was conducted in order to examine the applicability of the REMINT at the individual level, to explore its relationship with individual performance, in line with Hypothesis 4, and to validate a direct consensus version of the REMINT (see Study 3). Additionally, to assess incremental validity we included mental toughness, which has been shown in previous research to be a reliable predictor of performance (Gucciardi, Hanton, Gordin, Mallett, & Temby, 2015).

2.7.1 Method

2.7.1.1 Participants and Procedure.

We used a longitudinal design (students preparing for an oral exam in statistics) with four measurement points (t_1 to t_4). The first three measurement points were three weeks, two weeks, and one week before the oral exam, while the fourth was three days after the students had taken the exam. Online links to the questionnaires were sent via email. Participants received credit for their participation either in the form of participation points or the chance to win a €25 voucher for an online retailer.

One hundred sixty-two master students (mean age = 22.94, *SD* = 4.76, 80% female) who were enrolled in a statistical course (quantitative methods) took part in this study. To pass the course students were required to take a graded oral exam at the end of the semester. Random dropout (students who did not go on to fill out the questionnaires) led to a final total of 55 students who filled out all four questionnaires. G*Power (Faul et al., 2009) analysis verified that a sample size of 55 participants was sufficient (effect size of .4, resulting in an estimated power of .94).

2.7.1.2 Measures.

Reflection. We used the 32-item version of the REMINT (see Study 1) to test whether there were any differences between the individual and the team level. Items were framed as

individual-level items, and rated on a five-point Likert scale ranging from *strongly disagree* (1) to *strongly agree* (5).

Mental Toughness. Mental toughness was assessed using the eight-item mental toughness index (MT) developed by Gucciardi et al. (2015). Again, items were rated on a five-point Likert scale ranging from *strongly disagree* (1) to *strongly agree* (5).

Behavioural Criteria. During the weeks before the exam, students were asked when they started studying, how many hours they studied for each day, how often they met with others to study as a group, how often they went to tutorials and course lectures, and how often they initiated discussions about course-related content with others in forums or social networks.

Performance. Students' performance was measured by the grade he/she received after the oral exam. The course used a five-point grading system ranging from *very good* (1) to *deficient, course failed* (5). Additionally, grades were assigned using fractional values (e.g., 1.0, 1.3, 1.7 and so on; $M = 2.45$, $SD = 1.02$, Range = 4). To facilitate interpretation, performance scores were inverse-coded, so that high values represented high performance. The grades were assigned by two examiners, who discussed the performance of each student and reached an agreement afterwards.

2.7.2 Results and Discussion

We conducted a MLR CFA (using the expanded sample with $n = 162$) to examine the fit of the factorial models. As in Study 1, a four-factor model with no general reflection factor showed the best fit (for comparison, see Table 2.2). All further analysis was therefore conducted using the reduced sixteen-item REMINT. Correlations are reported in Table 2.5. Reliabilities ranged between .64 and .86 (see Table 2.5). As predicted in Hypothesis 4, most of the sub-dimensions of the REMINT on t_1 and t_3 were significantly and positively correlated to performance.

Table 2.5: Correlations and Descriptive Statistics for Study Variables in the Individual Sample

	M (SD)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Time 1																
1 IS Qn	3.30 (.65)	**.64**														
2 IS Ql	2.95 (.76)	.68**	**.77**													
3 IE Qn	3.57 (.69)	.48**	.51**	**.80**												
4 IE Ql	3.10 (.83)	.51**	.69**	.70**	**.86**											
5 MT	3.58 (.60)	.00	.01	.21	.20	**.82**										
Time 2																
6 IS Qn	3.39 (.66)	.43**	.56**	.59**	.49**	.19	**.75**									
7 IS Ql	3.09 (.71)	.40**	.53**	.44**	.55**	.18	.74**	**.80**								
8 IE Qn	3.59 (.67)	.34**	.37**	.64**	.50**	.22	.61**	.51**	**.78**							
9 IE Ql	3.29 (.74)	.26*	.44**	.54**	.53**	.25*	.63**	.66**	.68**	**.84**						
10 MT	3.56 (.56)	.12	.14	.24*	.16	.85**	.25*	.27*	.21	.24*	**.80**					
Time 3																
11 IS Qn	3.43 (.66)	.18	.32**	.48**	.40**	.02	.60**	.50**	.57**	.51**	.06	**.74**				
12 IS Ql	3.11 (.76)	.20	.40**	.47**	.52**	.18	.57**	.59**	.47**	.52**	.21	.78**	**.85**			
13 IE Qn	3.63 (.68)	.11	.20	.53**	.39**	.10	.38**	.36**	.58**	.48**	.13	.74**	.65**	**.85**		
14 IE Ql	3.29 (.70)	.05	.20	.39**	.41**	.18	.39**	.50**	.44**	.62**	.23	.63**	.73**	.70**	**.86**	
15 MT	3.64 (.64)	.08	.10	.29*	.21	.76**	.20	.25*	.18	.22	.84**	.02	.16	.16	.20	**.77**
Time 4																
16 Perf	2.45 (1.02)	.31*	.17	.44**	.28*	.30*	.19	.13	.30*	.14	.37**	.24	.28*	.33*	.41**	.32*

Note. *N* is between 77 (Time 1) and 55 (Time 4). IS = Information seeking. IE = Information evaluation. Qn = Quantity. Ql = Quality. MT = Mental toughness. Perf = Performance. Reliability alphas on the diagonal (bold).

*$p < .05$. **$p < .01$ (two-tailed).

Interestingly, on t_2 only the quantity of information evaluation and performance were significantly correlated. Mean scores on each dimension of the REMINT increased between t_1 to t_3.

A hierarchical regression analysis[2] (Table 2.6) revealed that the quantitative level of information-seeking at t_1 and the qualitative level of information evaluation at t_3 explained 21% of the variance of performance (adjusted R^2). Both variables significantly and individually predict performance, providing further support for Hypothesis 4. Furthermore, the results do not alter when the level of MT is accounted for. Power analysis underlines the significance of all three constructs (power ranges from .97 to 1.00). As further support for the validity of the REMINT, most of the behavioural criteria were significantly correlated to at least one plausible dimension of the REMINT during t_1 (see Appendix A).

Table 2.6: Hierarchical Regression of Individual Performance

	Performance (Grade)		
Variable	Step 1	Step 2	Step 3
IS quantity t_1	.31*	.28*	.30*
IE quality t_3		.38**	.34**
MT t_3			.29*
Adjusted R^2	.08	.21	.28
R^2	.10	.24	.32
Change R^2	.10*	.14**	.08*

Note. N = 55. Regression coefficients are standardized. IS = Information seeking. IE = Information evaluation. MT = Mental toughness.
*p < .05. **p < .01.

[2] When we included gender, age, or semester as control variables in the regression analyses, our pattern of results remained unchanged. We therefore did not include these in the further analysis.

Participants who reported higher quantitative information-seeking, for example, also reported significantly more meetings with study groups ($p < .01$), went to more tutorials ($p < .05$), and initiated more discussions on online platforms ($p < .01$). Participants reporting higher quantitative information evaluation values also reported that they started studying significantly earlier ($p < .01$) and spent longer studying each day ($p < .01$). Thus, we did not find significant correlations to these criteria during t_2 or t_3.

Overall, this study provides evidence of the criterion-related validity of the REMINT at the individual level and of our conception of reflection as a multi-factor construct. Results revealed the importance of the different dimensions, highlighting different learning phases. It seems to be more effective to focus initially on gathering as much information as possible in order to assess the current understanding of the learning material. Thus, reflecting on and evaluating (and perhaps adjusting) the way one learns becomes more effective during the later stages of learning. This finding is confirmed by (a) the significant correlations between the REMINT dimensions and the behavioural criteria during t_1, (b) the increasing strength of the correlations between the REMINT and performance, and (c) the increase in the mean values of the four REMINT dimensions over time. Thus, this particular result might be influenced by the nature of the task (i.e., studying for an exam). Examining different tasks with the REMINT might show different combinations of quality or quantity of information seeking and information evaluation to be more effective.

2.8 General Discussion

In this research, we systematically developed and validated a new multidimensional measure of reflection (i.e., the REMINT) across four studies and demonstrated that it has shown internal consistency and construct validity. Given that these results are derived from diverse samples, the findings suggest that the REMINT dimensions are robust across domains and contexts. Furthermore, the use of objective performance and behavioural criteria, as well

as longitudinal designs in Studies 3 and 4, strongly supports our hypothesis that the REMINT predicts performance independently of the composition methods (i.e., direct consensus composition or referent-shift consensus composition).

The current studies extend previous theory and research on team reflexivity in three meaningful ways. First, the REMINT distinguishes between two dimensions of psychological processes, namely quality and quality. For certain tasks (i.e., when the goal is to be as accurate as possible) quality and quantity aspects may be inversely related to one another, since quantity often demands that people launch straight into tasks and work as quickly as possible, and they may thus tend to pay less attention to accuracy and possible errors. Previous measures of reflection were concentrated mainly on quantity, thereby underestimating (a) the importance of qualitative aspects of reflection and (b) effects of possible trade-offs between focussing on quantitative and qualitative aspects of reflection. This might explain the inconclusive results. Konradt et al. (2016), for instance, argued that a team that is pressed for time might focus on superficial reflection, failing to reach real depth (i.e., accuracy), and making hasty decisions (i.e., speed). The REMINT allows us to examine the differential effects of the quantity and quality of reflection, including effects of different quantitative and qualitative ratios across time.

As second main extension of the previous reflexivity research, we put forward a conceptual framework in which reflexivity is composed of four interacting processes. Within this framework, we disentangled the key transition processes (i.e., information seeking and information evaluation), which are theoretically and conceptually related. We thereby provide a better and more accurate depiction of reflexivity and, at a more fundamental level, provide the basis for developing more elaborated theory and hypotheses. For instance, types and patterns of reflexive processes might now be delineated that have not previously been part of reflexivity research, permitting a continuum of forms, and resulting in a range of different

outcomes (bad, benign or beneficial). More specifically, our longitudinal analyses of individual reflection suggest that most effective individuals adjust their reflexive focus (i.e., focus on the quality or quantity of information seeking or evaluating) dynamically, according to situational demands. We expect these dynamic shifts to be applicable to the team level as well. These adjustments provide valuable information, and future research should therefore extend the quality/quantity differentiation to the transition processes of planning (Marks et al., 2001) and the action process of implementation, and examine their impact on performance.

Moreover, we validated a team and individual frame of reference version of the REMINT, providing a first step towards future multi-level frameworks for team reflexivity that identifies individual differences between team members at various points in time, including consensus and diversity, and the effect on action processes and performance (e.g., Mohammed & Harrison, 2013). Using a multi-level, multi-theory and multi-period framework (see Humphrey & Aime, 2014) should ultimately allow us to arrive at a more accurate description of team reflexivity and its effects on various outcomes.

One final way in which we extend previous reflexivity research is that is that our measure allows a better understanding not only of the team processes involved in reflexivity, but also of how to train team leaders, mentors or coaches who can guide and facilitate the reflexive process. Facilitators could be trained to identify critical situations, so that they can help teams to adjust their reflexive focus to ensure that it fits the particular situation they are encountering. Future research should therefore match specific aspects of the work context (e.g., task characteristics, level of task completion, or task environments) to specific foci on reflexive behaviours in order to ascertain the optimal fit.

2.9 Limitations

There are five limitations to our findings that may offer useful areas for future research. First, the use of quasi-teams in Studies 1 and 2, and the small sample size in Study

3, limits the external validity and the robustness of the inferences to be drawn from both studies. The small sample size would, however, also indicate a conservative test of our hypotheses, as we lack the statistical power to detect potentially less obvious relationships. Nevertheless, most of the results found were consistent with our hypothesis and provided valuable information on the validity of the REMINT and its applicability in different contexts.

It may also seem to be a limitation of this research that we obtained low reliabilities for the quantitative scales, in particular the quantitative information-seeking subscale, where values occasionally fell below common thresholds (Nunnally & Bernstein, 1994). However, as Thorsen and Bjorner (2010) have pointed out, low reliabilities do not always lead to biased estimates, because adequate reliability depends on the study and the purpose. For instance, low reliability will not bias the estimates for scales that are used as endpoint variables. Moreover, detrimental effects of low reliabilities of independent variables may be countered by applying structural equation modelling that take measurement errors into account. We assume that the low reliabilities might be attributed to the fact that the subscales are comparatively short, given the conceptual breath of each construct. Future research should investigate this issue further, using larger samples.

The third limitation concerns the general challenge of using questionnaires and other research methods, which rely chiefly on language to make phenomenological conclusions. As noted by Alvesson and Kärreman (2000), the linguistic turn in organizational research using language as a mirror for social reality may lead to inconsistent findings as a result of "questionnaire items forcing respondents to reply in a highly-constrained way" (p. 140). Thus, these statistical data may be more indicative of the way the language of questionnaires is interpreted by the respondents and be less valid for the research itself for the research issue itself. Future research should focus much more on specific core situations of reflexivity. According to our questionnaire, it seems important to provide clear introductions for the items

which are based on such core situation (e.g., the following items are about…) so that teams really understand how to differentiate between quality and quantity. Future research should therefore look at ways of providing clearer introductions for both subscales in order to overcome the linguistic concerns and enable valid conclusions to be drawn.

The fourth limitation concerns the possibility of a general halo effect in the teams in Study 3. Teams that had a positive climate may have rated all aspects of their functioning more positively (including reflection), regardless of their actual levels of reflection. Because these teams might have been for example more cohesive and optimistic, they may have ended up performing well. Since we did not measure general affect in Study 3, this possibility cannot be ruled out (cf. Barsade & Knight, 2015).

Finally, despite the promising results from our measure, the reflexivity framework needs further theorizing and empirical validation. In capturing two key transition processes of reflection (i.e., information seeking and information evaluation), other relevant transition and action processes in teams are planning and implementation, which have been shown to explain additional variance in team performance (Konradt et al., 2015; Schippers et al., 2007; Wiedow & Konradt, 2011). We therefore point out the importance of assessing the complete reflexivity framework in future research, and the extension of the quality and quantity differentiation on the reflexive processes of planning and implementation. Moreover, it is also not yet clear whether our findings are equally valid for all types of teams (c.f. Sundstrom, 1999). Likewise, different types of teams might exhibit different patterns of reflection, because they are bound to different timeframes (time-limited vs. permanent and ongoing) or work on different tasks (e.g., laboratory vs. field). Consequently, future research should seek to expand on the current findings by assessing and relating information-seeking and information evaluation processes to the planning and implementation carried out by different types of teams working on different tasks.

2.10 References

Alvesson, M., & Kärreman, D. (2000). Varieties of discourse: On the study of organizations through discourse analysis. *Human Relations, 53*, 1125–1149. doi:10.1177/0018726700539002

Anseel, F., Beatty, A. S., Shen, W., Lievens, F., & Sacket, P. R. (2015). How are we doing after 30 years? A meta-analytic review of the antecedents and outcomes of feedback-seeking behavior. *Journal of Management, 41*, 318–348. doi:10.1177/0149206313484521

Anseel, F., Lievens, F., & Schollaert, E. (2009). Reflection as a strategy to enhance task performance after feedback. *Organizational Behavior and Human Decision Processes, 110*, 23–35. doi:10.1016/j.obhdp.2009.05.003

Barsade, S. G., & Knight, A. P. (2015). Group affect. *Annual Review of Organizational Psychology and Organizational Behavior, 2*, 21–46. doi:10.1146/annurev-orgpsych-032414-111316

Brahm, T. (2009). *Entwicklung von Teamkompetenz durch computergestützte kollaborative Lernprozesse: Eine empirische Analyse [Development of team competence by means of computer-aided collaborative learning processes: An empirical analysis]*. (Dissertation), University of St. Gallen, St. Gallen.

Brown, T. A. (2015). *Confirmatory Factor Analysis for Applied Research* (2nd ed.). New York, NY: Guilford Publications.

Bryant, F. B., & Satorra, A. (2012). Principles and practice of scaled difference chi-square testing. *Structural Equation Modeling: A Multidisciplinary Journal, 19*, 372–398. doi:10.1080/10705511.2012.687671

Bulmer, M. (1979). Concepts in the analysis of qualitative data. *Sociological Review, 27*, 651–677. doi:10.1111/j.1467-954X.1979.tb00354.x

Carter, S. M., & West, M. A. (1998). Reflexivity, effectiveness, and mental health in BBC-TV production teams. *Small Group Research, 29,* 583–601. doi:10.1177/1046496498295003

Chan, D. (1998). Functional relations among constructs in the same content domain at different levels of analysis: A typology of composition models. *Journal of Applied Psychology, 83,* 234–246. doi:10.1037//0021-9010.83.2.234

Churchill, G. A. (1979). A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research, 16,* 64–73. doi:10.2307/3150876

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (3rd ed.). Mahway, NJ: Lawrence Erlbaum Associates.

Comrey, A. L., & Lee, H. B. (1992). *A First Course in Factor Analysis*. Hillsdale, NJ: Erlbaum.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281–302.

De Dreu, C. K. W. (2007). Cooperative outcome interdependence, task reflexivity, and team effectiveness: A motivated information processing perspective. *Journal of Applied Psychology, 92,* 628–638. doi:10.1037/0021-9010.92.3.628

De Jong, B. A., & Elfring, T. (2010). How does trust affect the performance of ongoing teams? The mediating role of reflexivity, monitoring, and effort. *Academy of Management Journal, 53,* 535–549. doi:10.5465/AMJ.2010.51468649

DeVellis, R. F. (2003). *Scale development: theory and applications* (2nd ed.). Thousand Oaks, CA: Sage.

Edmondson, A. (1999). Psychological safety and learning behaviour in work teams. *Administrative Science Quarterly, 44,* 350–383. doi:10.2307/2666999

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41,* 1149–1160. doi:10.3758/BRM.41.4.1149

Fisher, R. P., & Geiselman, R. E. (1992). *Memory Enhancing Techniques for Investigative Interviewing: The Cognitive Interview*. Springfield, IL: Charles C. Thoma.

Gabelica, C., Van den Bossche, P., De Maeyer, S., Segers, M., & Gijselaers, W. (2014). The effect of team feedback and guided reflexivity on team performance change. *Learning and Instruction, 34,* 86–96. doi:10.1016/j.learninstruc.2014.09.001

Gabelica, C., Van den Bossche, P., Segers, M., & Gijselaers, W. (2014). Dynamics of team reflexivity after feedback. *Frontline Learning Research, 2,* 64–91. doi:10.14786/flr.v2i2.79

Gevers, J. M. P., van Eerde, W., & Rutte, C. G. (2001). Time pressure, potency, and progress in project groups. *European Journal of Work and Organizational Psychology, 10,* 205-221. doi:10.1080/13594320143000636

Gucciardi, D. F., Hanton, S., Gordon, S., Mallett, C. J., & Temby, P. (2015). The concept of mental toughness: Tests of dimensionality, nomological network, and traitness. *Journal of Personality*, *83*, 26–44. doi:10.1111/jopy.12079

Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods*, *1*, 104–121. doi:10.1177/109442819800100106

Hoelter, J. W. (1983). The analysis of covariance structures: Goodness-of-fit indices. *Sociological Methods & Research, 11,* 325–344. doi:10.1177/0049124183011003003

Humphrey, S. E., & Aime, F. (2014). Team microdynamics: Toward an organizing approach to teamwork. *The Academy of Management Annals, 8,* 443–503. doi:10.1080/19416520.2014.904140

Ilgen, D. R., Hollenbeck, J. R., Johnson, M., & Jundt, D. (2005). Teams in organizations: from input-process-output models to imoi models. *Annual Review of Psychology, 56*, 517–543. doi:10.1146/annurev.psych.56.091103.070250

Konradt, U., & Eckardt, G. (2016). Short-term and long-term relations among reflection and performance in teams: Evidence from a four-wave longitudinal study. *European Journal of Work and Organizational Psychology, 25, 804–818*. doi:10.1080/1359432X.2016.1160058

Konradt, U., Otte, K.-P., Schippers, M. C., & Steenfatt, C. (2016). Reflexivity in teams – A review and new perspectives. *The Journal of Psychology: Interdisciplinary and Applied, 150*, 153–174. doi:10.1080/00223980.2015.1050977

Konradt, U., Schippers, M. C., Garbers, Y., & Steenfatt, C. (2015). Effects of guided reflexivity and team feedback on team performance improvement: The role of team regulatory processes and cognitive emergent states. *European Journal of Work and Organizational Psychology, 24*, 777–795. doi:10.1080/1359432X.2015.1005608

Lee, I. A., & Preacher, K. J. (2013). Calculation for the test of the difference between two dependent correlations with one variable in common [Computer software]. Retrieved from http://quantpsy.org

Lindell, M. K., Brandt, C. J., & Whitney, D. J. (1999). A revised index of interrater agreement for multi-item ratings of a single target. *Applied Psychological Measurement, 23*, 127–135. doi:10.1177/01466219922031257

MacCallum, R. (1986). Specification searches in covariance structure modelling. *Psychological Bulletin, 100*, 107–120. doi:10.1037/0033-2909.100.1.107

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4*, 84–99. doi:10.1037/1082-989X.4.1.84

Marks, M. A., Mathieu, J. E., & Zaccaro, S. J. (2001). A temporally based framework and

taxonomy of team processes. *Academy of Management Review, 26*, 356–376. doi:10.5465/AMR.2001.4845785

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*, 741–749. doi:10.1037/0003-066x.50.9.741

Mohammed, S., & Harrison, D. A. (2013). The clocks that time us are not the same: A theory of temporal diversity, task characteristics, and performance in teams. *Organizational Behavior and Human Decision Processes, 122*, 244–256. doi:10.1016/j.obhdp.2013.08.004

Moreland, R. L., & McMinn, J. G. (2010). Group reflexivity and performance. *Advances in Group Processes, 27*, 63–95. doi:10.1108/s0882-6145(2010)0000027006

Muthén, B. & Asparouhov, T. (2012). Bayesian SEM: A more flexible representation of substantive theory. *Psychological Methods*, 17, 313-335. doi:10.1037/a0026802

Muthén, L. K., & Muthén, B. O. (1998–2016). *Mplus User's Guide. Seventh Edition*. Los Angeles, CA: Muthén & Muthén.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric Theory* (3rd ed.). New York, NY: McGraw-Hill.

Rousseau, D. (1985). Issues of level in organizational research: Multi-level and cross-level perspectives. In L. Cummings & B. Saw (Eds.), *Research in Organizational Behaviour* (Vol. 7, pp. 1–37). Greenwich, CT: JAI.

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent Variables Analysis: Applications for Developmental Research* (pp. 399–419). Thousand Oaks, CA: Sage.

Satorra, A. (2000). Scaled and adjusted restricted tests in multisample analysis of moment

structures. In D. D. H. Heijmans, D. S. G. Pollock, & A. Satorra (Eds.*), Innovations in Multivariate Statistical Analysis: A Festschrift for Heinz Neudecker* (pp. 233–247). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Savelsbergh, C., Van der Heijden, B. I. J. M., & Poell. R. F. (2009). The development and empirical validation of a multi-dimensional measurement instrument for team learning behaviors. *Small Group Research, 40*, 578–607. doi:10.1177/1046496409340055

Schippers, M. C., Den Hartog, D. N., & Koopman, P. L. (2007). Reflexivity in teams: A measure and correlates. *Applied Psychology: An International Review, 56*, 189–211. doi:10.1111/j.1464-0597.2006.00250.x

Schippers, M. C., Edmondson, A. C., & West, M. A. (2014). Team reflexivity as an antidote to team information-processing failures. *Small Group Research, 56*, 731–769. doi:10.1177/1046496414553473

Schippers, M. C., Edmondson, A. C., & West, M. A. (in press). Team reflexivity. In J. M. Levine & L. Argote (Eds.), *Handbook of Group and Organizational Learning*. Oxford: Oxford University Press.

Schippers, M. C., Homan, A. C., & van Knippenberg, D. (2013). To reflect or not to reflect: Prior team performance as a boundary condition of the effects of reflexivity on learning and final team performance. *Journal of Organizational Behavior, 34*, 6–23. doi:10.1002/job.1784

Schippers, M. C., West, M. A., & Edmondson, A. C. (in press). Team reflexivity and innovation. In R. Rico, N. M. Ashkanasy, & O. B. Salas (Eds.) *Handbook of the Psychology of Teamwork and Collaborative Processes*. New Jersey: Wiley-Blackwell.

Smith-Crowe, K., Burke, M. J., Cohen, A., & Doveh, E. (2014). Statistical significance criteria for the rWG and average deviation interrater agreement indices. *Journal of Applied Psychology, 99*, 239–261. doi:10.1037/a9934556.

Song, X. Y. and Lee, S. Y. (2012). *Basic and Advanced Bayesian Structural Equation Modeling: With Applications in the Medical and Behavioral Sciences*. London: Wiley.

Sundstrom, E. (1999). The challenges of supporting work team effectiveness. In E. Sundstrom & Associates (Ed.), *Supporting work team effectiveness* (pp. 3–23). San Francisco: Jossey-Bass.

Thorsen, S. V., & Bjorner, J. B. (2010). Reliability of the Copenhagen psychosocial questionnaire. *Scandinavian Journal of Public Health, 38, 25–32. doi:10.1177/1403494809349859*

Van Dick, R., & West, M. A. (2005). *Teamwork, Teamdiagnose, Teamentwicklung: Praxis der Personalpsychologie. [Teamwork, Team diagnosis, Team development*. Practices in Personnel psychology]. Göttingen: Hogrefe.

van Mierlo, H., Vermunt, J. K., & Rutte, C. G. (2009). Composing group-level constructs from individual-level survey data. *Organizational Research Methods, 12*, 368–392. doi:0.1177/1094428107309322

West, M. A. (2000). Reflexivity, revolution, and innovation in work teams. In M. M. Beyerlein, D. Johnson, & S. T. Beyerlein (Eds.), *Product Development Teams* (Vol. 5, pp. 1–29). Stamford, Connecticut: JAI Press.

West, S. G., Taylor, A. B., & Wu, W. (2012). Model fit and model selection in structural equation modelling. In R. H. Hoyle (Ed.), *Handbook of Structural Equation Modelling* (pp. 209-231). New York: Guilford.

Wiedow, A., & Konradt, U. (2011). Two-dimensional structure of team process improvement: Team reflection and team adaptation. *Small Group Research, 42*, 32–54. doi:10.1177/1046496410377358

Zyphur, M. J. & Oswald, F. L. (2015). Bayesian estimation and inference: A user's guide. *Journal of Management, 41*, 390–420. doi:10.1177/0149206313501200

Kapitel 3: Studie 2

Disentangling the Relationship between Team Reflection and Team Performance: The Role of Quality and Quantity

Otte, K.-P., Konradt, U., & Oldeweme, M. (2017). *Disentangling the relationship between team reflection and team performance: the role of quality and quantity*. Unpublished manuscript.

Abstract

In spite of a rising number of team reflexivity studies, previous research has mainly focused on the quantity of team reflexive processes, and ignored their quality. Moreover, the measures used to assess team reflection, were mainly based on self-ratings, which might have been distorted by biases. In a laboratory setting using 47 three-person teams, we therefore explored the relationship between the qualitative and quantitative team reflection dimensions, and team performance improvement, and compared the validity of self- and other-ratings of both dimensions. We expected a negative relationship between quantitative team reflection and performance improvement, and a positive relationship between qualitative team reflection and performance improvement. Additionally, we hypothesized that feedback given by the preceding team performance is negatively related to subsequent team reflection. Finally, we examined that the action-phase process of implementation mediates the team reflection – performance improvement relationship. The results support most of our hypotheses, provide relevant insights into the often-ambiguous relationship between team reflection and team performance, and point towards the validity of self-rated team reflection measure. Implications of our findings for theories, research, and practices are discussed.

3.1 Introduction

During the last years, team reflexivity has emerged as important factor in team functioning (Konradt, Otte, Schippers, & Steenfatt, 2016). Team reflexivity refers to the "extent to which team members collectively reflect upon the team's objectives, strategies, and processes, as well as their wider organizations and environments, and adapt them accordingly" (West, 2000, p. 3). Reflexive teams are generally assumed to show higher levels of performance (Konradt & Eckardt, 2016) and innovation (Schippers, West, & Dawson, 2015). Although previous research has constantly contributed to the progress in understanding the effects of team reflection on team performance (e.g., Schippers, Homan, & van Knippenberg, 2013; Konradt & Eckardt, 2016), this study addresses two critical gaps in the pertaining literature.

The first gap in previous research is the conception of team reflexivity as omnibus concept (Otte, Konradt, Garbers, and Schippers, 2017), that integrated multiple team processes (i.e., reflection, planning, and implementation) within a single construct. Moreover, previous measures only captured how extensively teams reflected (i.e., quantity), but ignored important aspects of depth and detail (i.e., quality). Consequently, this might have led to ambiguous or inconclusive results (Konradt et al., 2016). In particular, as time is often a scarce resource, reflecting upon too many different topics could keep teams from reaching substantial depth. Depth, however, can be critical for developing adequate solutions for previously unsolved problems, whereas extensive width might hinder their development. Hence, the current research is the first that examines the diverging effects of qualitative and quantitative team reflection on team performance.

Secondly, team reflection research mainly focused on data obtained with self-reports via questionnaires. Self-reports are, however, often distorted by biases, as for example halo or recency effects (Podsakoff & Organ, 1986), and are sometimes assumed to hinder a more

detailed examination of team reflection processes. A possible solution to address the validity problem is the usage of other-reports. Yet, no previous study has examined the degree to which self- and other-rated team reflection measures overlap. Therefore, we compared questionnaire-based self-reports and non-obtrusive behavioral-based other-reports, and determine the degree to which both measures are correlated.

3.2 Validity of Team Reflection

Psychological and organizational research relies to a large extent on questionnaire data generated through self-report measures. Self-report measures are, however, associated with a number of problems, including common method biases (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003), social desirability (Berry, Carpenter, & Barratt, 2012), or consistency seeking and self-presentation (Paulhus & Vazire, 2007). Consequently, the validity of self-report measures of team reflection may be restricted. These restrictions become apparent when for example comparing self- and other-rated team performance. In a field study on 50 work teams, Wiedow and Konradt reported a positive relation between team reflection and self-rated coordination success, as well as self-rated team performance, but a negative relation between reflection and supervisor-rated coordination success, as well as supervisor-rated team performance. These results indicate substantial differences between one owns perceptions and the perceptions of others.

A possible solution to overcome the validity concerns is to draw on behavioral data and content analysis (Rourke & Anderson, 2004). Although West (2000) and more recently Schippers, Edmondson, and West (2014) called for content analyses of team reflection, to our knowledge, only two team reflexivity studies have used content analysis for examining team reflection. Gurtner, Tschan, Semmer, and Nägele (2007) examined 49 teams performing a laboratory military air-surveillance task in three different conditions (individual reflection, team reflection, and control condition). Drawing on the team communication protocols, they

demonstrated that individual reflection was more effective than team reflection. They also found that teams more often focused on general, rather than task-specific discussions, which explained the inferior performance for team reflection. Gabelica, Van den Bossche, Segers, and Gijselaers (2014), on the other hand, examined in a laboratory setting whether teams engaged in full reflexivity cycles (i.e., reflecting, planning, and implementing, cf. West, 2000; Schippers, Den Hartog, & Koopman, 2007). Using verbal protocols and content analysis, they found that teams never completed a full cycle, and often returned to previous steps within one cycle.

However, in spite of the additional value provided by behavioral data, important questions have remained unaddressed. Most importantly, it is yet unclear to which degree self- and other rated team reflection data overlap, and hence, whether self-report team reflection measures are indeed lacking validity that can be provided by other-reports. In particular, the recent conceptualization of team reflection with a quantitative and qualitative dimension (Otte et al., 2017) may pose additional difficulties for participants to make accurate assessments of their reflection. Moreover, as both types of measurement possibly diverge, results from studies that used questionnaires might differ from studies using other-ratings based on behavioral data.

We therefore argue that it is indispensable to investigate the degree to which both measures are correlated. In the empirical research, the strength of the relations between self- and other-rated measures can be generally judged as moderate (cf. Cohen, 1988), as for instance in a meta-analytical examinations of counterproductive work behaviors (Berry et al., 2012), organizational citizenship behavior (Carpenter, Berry, & Houston, 2014), or self- and other-rated job performance (Heidemeier & Moser, 2009). Hence, as these results are extracted from a large number of different studies, we also expect a moderate relation between self- and other-rated quantitative team reflection. Thus, our first hypothesis is:

Hypothesis 1a: Self-rated and other-rated quantitative reflection measures are moderately positively related.

However, the relation between self-rated and other-rated qualitative team reflection could be more complex. For instance, we consider the simple situation with two reflexive utterances within a team. The first utterance was qualitatively on a low level, whereas the second utterance was on a qualitatively high level, resulting in an overall medium quality of team reflection. Thus, in order to increase the quality, one of the team members is required to make another high-quality utterance. However, this will also automatically increase the quantity of the discussion. As a result, both measures will be highly correlated, which is not atypical, as Otte et al. (2017) for instance already observed relatively high correlations between qualitative and quantitative team reflection assessed with questionnaires. However, since we only expect a moderate correlation for the quantitative dimensions, both qualitative dimensions will probably be impacted by diverging forms of quantitative reflection. Hence, in order to determine the relation between self- and other-rated qualitative team reflection, we argue that it has to be controlled for quantitative reflection.

Hypothesis 1b: Self-rated qualitative team reflection, controlled for self-rated quantitative team reflection, is positively related to other-rated qualitative team reflection, controlled for other-rated quantitative team reflection.

Finally, based on the expected moderate correlations, we explore whether other-rated team reflection data can explain additional variance above self-rated team reflection.

Research question: Can other-rated team reflection explain variance above self-rated team reflection?

3.3 Feedback and Team Reflection

A shortcoming of previous team reflection research is that little attention has been directed towards temporal issues. Based on the conception of input-mediator-output-input

models (IMOI; Ilgen, Hollenbeck, Johnson, & Jundt, 2005), and the recurring phase model of team processes (Marks, Mathieu, & Zaccaro, 2001), Konradt et al. (2016) conceived reflection as a team regulatory process (DeShon, Kozlowski, Schmidt, Milner, & Wiechmann, 2004) that takes place during performance episodes. Performance episodes are "distinguishable periods of time over which performance accrues and feedback is available" (Marks et al., 2001, p. 359). Moreover, performance episodes consist out of transition-phase and action-phase processes through which teams regulate their functioning and, consequently, goal attainment (Marks et al., 2001).

Nevertheless, self-regulation in the form of purposefully directing reflexive processes, requires the availability of feedback (Konradt et al., 2016). Feedback can be conceived as "information about the actual performance or actions of a system used to control the future actions of a system" (Nalder, 1979, p. 310). Multiple studies have highlighted the significance of feedback in the reflection context. For instance, Gabelica, Van den Bossche, De Maeyer, Segers, and Gijselaers (2014) in a laboratory simulation study with one hundred and six dyads, and Konradt, Schippers, Garbers, and Steenfatt (2015) in a laboratory decision making task with 98 student teams, showed that the combination of feedback and reflection was more effective than feedback or reflection alone. Yet, these studies exclusively relied on a discrete feedback manipulation in the form of a 'feedback' and a 'no feedback' condition. Organizational teams, however, often receive feedback in the form of preceding team performance (e.g., sales), which is more fine grained. Hence, as "team behavior regulation is primarily focused on reducing discrepancies between given and actually attained goals" (Konradt et al., 2016, p. 164), a large discrepancy between a given and an actually attained goal should result in more effort to reduce the discrepancy (Carver & Scheier, 1998). In contrast, a small discrepancy should result in less effort, because the team perceives its current performance as sufficient. Until now, only one study has examined this relation. Using

a four-wave longitudinal study with 97 teams working on a business simulation task, Konradt and Eckardt (2016) examined whether feedback operationalized as previous performance, affected subsequent reflection. Although they did not find the expected direct effect of team performance on subsequent team reflection, a declining team reflection trajectory and an increasing team performance trajectory pointed towards a generally negative relation between both variables. Hence, to further validate this finding, we expect a negative relation between feedback and subsequent team reflection.

Hypothesis 2: Positive/Negative feedback in terms of positive/negative team performance is related to a reduction/increase in subsequent qualitative and quantitative team reflection (both for self-rated and other-rated).

3.4 Team Reflection and Performance

Despite the large number of studies examining the relationships between reflection and outcomes, as for example performance, Konradt et al. (2016) concluded in their review that results have been mixed and ambiguous (e.g., De Dreu, 2002; Wiedow & Konradt, 2011). Konradt and colleagues therefore proposed that these results might stem from the understanding of team reflection as a quantitative construct, which captures only how *often* a team reflects (i.e., frequency, see also Moreland and McMinn, 2010; Schippers et al, 2014). This understanding ignored important qualitative elements in reflection that capture its depth and detail. For instance, teams that reflect upon their previous strategies and processes on a very superficial level, might miss more fundamental aspects that negatively impacted team performance, even though they spent a lot of time reflecting.

Consequently, an important element in reflection might not be whether teams reflect, but the particular reflexive behaviors they focus upon. Notably, the effects on outcomes might also be contingent on external factors. Schippers et al. (2015) for example found in a field study with 98 primary health care teams, that the effects of reflection on performance were

contingent on the quality of the work environments. Similarly, De Dreu (2002) showed that team reflection only had a positive impact on team innovation and team effectiveness, when minority dissent was high (i.e., when a minority in a group opposes the thinking of the majority), and Schippers et al. (2013) only found a positive relation between team reflection and performance, when the initial performance of a team was poor.

However, on a more general level, an important contingency might be temporal constraints. As time is often a scarce resource, teams have to manage their activities by differentiating between relevant and irrelevant processes (cf. resource allocation theory, Norman & Bobrow, 1975), and shifting their resources accordingly. Thus, one reflexive behavior might come at the expense of the other. Similar to the effects of speed-accuracy trade-offs in decision making tasks (Förster, Higgins, & Bianco, 2003), the effects of team reflection under temporal constraints may be contingent on a quantity-quality trade-off. Reflecting upon too many topics could hinder teams from reaching substantial depth and result in incomplete or incorrect conclusions that affect consecutive processes and outcomes. Quality, on the other hand, might be a vital component in team reflection, since it enables teams to elaborate selected topics in more detail (Otte et al., 2017), which should positively affect team performance. Moreover, Konradt et al. (2015), as well as Schippers et al. (2013), argued that the change in performance (i.e., performance improvement) is a better performance indicator, as it gives a clearer picture of the possible effects of team reflection on performance. We therefore hypothesize:

Hypothesis 3: Under time-limited conditions, self-rated/other-rated quantitative team reflection has a detrimental effect on team performance improvement, when qualitative team reflection remains unchanged.

Hypothesis 4: Under time-limited conditions self-rated/other-rated qualitative team reflection positively affects team performance improvement, when quantitative team reflection remains unchanged.

3.5 Reflection, Implementation, and Performance

A central element in team reflexivity research is the mediating role of implementation (e.g., Konradt et al., 2015; Schippers et al., 2007; West, 2000). West (2000) defined implementation (originally named adaptation) as "goal-directed behaviors relevant to achieving the desired changes in team objectives, strategies, processes, organizations, or environments" (p. 6). Konradt et al. (2016) furthermore conceived implementation as a relevant team-regulatory action-phase process that mediates the relation between team reflection and team performance. However, the exact measurement of implementation might be difficult to achieve, because implementation strongly depend on contexts and task types. For instance, Gurtner et al. (2007) calculated a relatively complex implementation score that captured how the teams changed all task-relevant behaviors from one round to the subsequent round. Nevertheless, this approach may not be always feasible, due to the inaccessibility of objective implementation indicators, or the complexity of the organizational system in which a team operates. Hence, to reduce the complexity of the construct, Schippers et al. (2007) proposed to conceive implementation as "the extent to which teams live up to agreements" (p. 192). This method allows to measure implementation independently from the actual number of changes that were made between for example two team performance episodes.

Still, very few studies examined the mediating role of implementation between team reflection and team performance. Gurtner and colleagues (2007) found that implementation fully mediated the relationship between team reflection and team performance. Similarly, Konradt et al. (2015) examined implementation within a mediational model, in which the effects of team reflection were also fully mediated by team mental models and

implementation. In contrast, although not explicitly examined as mediator, Wiedow and Konradt (2011) provided evidence that team implementation explains unique variance above team reflection, but did not diminish the effects of team reflection.

Moreover, previous research that conceived implementation as living up to agreements often reported very high correlations between implementation and team reflection. For instance, Schippers et al. (2007) found high positive relationships ($r > 0.6$) between implementation and reflection in a study on 60 school management teams. A similar result was reported by Wiedow and Konradt (2011), and in the laboratory study from Konradt et al. (2015). These high correlations could possibly lead to an underestimation of the direct effect of team reflection on performance in a mediational model, because of range restrictions (i.e., floor and ceiling constraints) that render unique variance in the reflection-performance relationship inaccessible when including implementation (cf. Cole, Bedeian, Hirschfeld, & Vogel, 2011). Furthermore, the operationalization of implementation as living up to agreements does not capture the quality of the solution developed by the team. Hence, a team that developed and implemented a qualitatively high solution would not be distinguishable from a team that developed and implemented a qualitatively low solution. Consequently, relevant variance would be lost when not considering direct effects of team reflection on performance. Hence, we argue that implementation conceived as living up to agreements partially mediates the relation between team reflection and performance.

Hypothesis 5: Team implementation partially mediates the relation between team reflection and performance improvement.

Additionally, as qualitative team reflection should enable teams to conduct deeper and more elaborate evaluations of their performance, this should lead to the development of new, more feasible strategies that are likely to be implemented. Consequently, an increase in

qualitative team reflection, while quantitative team reflection stays the same, should positively affect implementation.

Hypothesis 6: Qualitative team reflection, controlled for quantitative team reflection, is positively related to implementation.

In contrast, an increase in quantitative team reflection, while qualitative team reflection stays the same, should negatively impact implementation, because the development of new strategies is based on rather superficial team reflection.

Hypothesis 7: Quantitative team reflection, controlled for qualitative team reflection, is negatively related to implementation.

3.6 Method

3.6.1 The Task

In the current study, we used a more complex version of the weather company task (Konradt et al., 2015; Wiedow, Konradt, Ellwart, & Steenfatt, 2013). The task is based on a hidden-profile paradigm (Lu, Yuan, & McLeod, 2012; Stasser & Titus, 1985) and requires teams of three members to collaboratively work in a fictional travel agency that handles client requests. Each client has several requests for his holiday, as for example going to places with high temperatures, or low precipitation. Overall, each client requested four out of nine different categories: (1) temperature, (2) precipitation, (3) wind speed, (4) number of other tourists, (5) number of bars, (6) number of restaurants, (7) cultural offerings, (8) number of recreational activities, and (9) level of crime. The first four of the categories varied according to the month the client wanted to travel (e.g., higher temperatures during the summer months), the other five did not change. Participants received the client request in written form. Each request consisted out the month of travel, and four of the nine aforementioned travel categories. Overall, the task consisted out of two consecutive rounds with eleven possible client requests per round.

The task of the teams was to find the best-fitting travel route for each client, by choosing one out of three available routes (e.g., a journey through Europe). Each route was divided into six cities. Team member had access to datasheets that contained information on the specific conditions in each of the city (i.e., temperature, number of bars, or crime rates). Thus, as the task was based on a hidden paradigm (Stasser & Titus, 1985), the presented information on the datasheets were incomplete and each team member had access to a small part of the information. Hence, the team members had to exchange the relevant information to determine the best fitting route.

In order to compare information from different data sources (e.g., wind speeds given in meters per seconds, and temperatures in degrees), information provided on the datasheets had to be converted from raw values to standardized value points, that ranged on a scale from one to four. For instance, if a client requested high temperatures, a high value point indicated a good fit for the respective city. Instructions on how to convert raw values to value points were accessible by the participants during the whole task.

After assigning a client to one of the three available routes, teams could ask for the next client. Participants were asked to complete as many clients requests as possible within one round, while working as accurately as possible. Participants were informed that the overall performance score would be calculated by 80% from the correctly calculated points and 20% from the number of clients. We chose this weighted value in order to encourage the participants to consider a wider range of different strategies, and reflect upon their effectiveness afterwards.

3.6.2 Participants and Procedures

Participants were recruited via email, notice boards, and direct contact. Overall, N = 141 students (82% from the psychology department, the 18% from other departments) in 47 three-person teams studying at a German university participated in the study. All participants

were informed that their conversations were audiotaped for later analysis. Due to connectivity problems with the internet, the questionnaire data from one group could not be saved. Moreover, audio data for another group were not recorded due to an empty battery. Therefore, overall sample size was reduced to $N = 46$ teams. 72.5% were female, and the average age was 22.91 years ($SD = 4.38$). Each trial was managed by two investigators, one for the primary contact with the participants, the other for processing the feedback and preparing the following round.

At the beginning of both rounds, participants received instructions for the respective round. During the round (i.e., action phase, t_0), they tried to finish as many client requests as possible. After the first round (t_1), teams received feedback on their performance and were given the opportunity to reflect, and answered reflection the questionnaires (see measures) afterwards. Subsequently, they started working on the second round (t_2)

Participants were allowed to reflect for a maximum of five minutes. Moreover, based on previous studies (e.g., Konradt et al., 2015), we provided reflection guidelines for each team, in order to avoid teams not reflecting at all. Implementation was measured directly after the second action phase. After completing both rounds, the teams were thanked for their participation and signed a non-disclosure agreement that instructed them not to talk to others about the contents of the task. Participants received rewards in the form of participation points. Furthermore, they had the chance to win money and vouchers if they finished among the best six teams.

3.6.3 Measures

Team Reflection. Quality and quantity of team reflection were measured with the team information evaluation dimensions of the REMINT (Reflection measure for individuals and teams, Otte et al., 2017). A sample item of qualitative team reflection was ‘we made very detailed evaluations of the quality of our work’, whereas a sample item of quantitative team

reflection was ‘we made very frequent assessments of how successful our procedures were’. The response format was a five-point Likert scale ranging from *strongly disagree* (1) to *strongly agree* (5). The participants were asked to base their answers on the team reflection phase.

Implementation. Implementation was measured with a three-item questionnaire based on the implementation measure developed by Schippers et al. (2007). A sample item was ‘in this team we implement changes we agreed upon’. The response format was a five-point Likert scale ranging from *strongly disagree* (1) to *strongly agree* (5).

Performance. As outlined above, performance was measured as the weighted average of the number of correctly calculated value-points (80%) and the number of correctly assigned routes to the handled clients (20%). To compare both measures, we transformed them into value-points per minute, and clients per minute.

3.6.4 Quantitative Content Analysis and Development of a Coding Scheme

Quantitative content analysis (QCA) is defined as “a research technique for the systematic, objective, and quantitative description of the manifest context of communication” (Berelson, 1952, p. 18). Rourke and Anderson (2004) refer to a process of segmenting communication content into units, which are then assigned to categories, which are in return used for further quantitative analysis, as for example the usual covariance-based methods.

In this study, we followed the guidelines outlined in Rourke and Anderson (2004) for developing an appropriate coding scheme for the team reflection data. The authors refer to five steps to develop a theoretically valid coding scheme. In the first step, the researcher should identify the purpose of the coding data. As indicated by our hypotheses, the purpose of the coded data is to explore the relation of coded team reflection data to questionnaire-based team reflection data. The second step involves the identification of behaviors that represent the construct. Rourke and Anderson elaborate that conducting content analysis without a

systematic identification process may result in “data being uninterpretable or rival interpretations that are more plausible than those offered by the content analysis” (p. 11). As the target-content was manifest (i.e., team discussion-based elements in team reflection, Rourke and Anders, 2004), we concentrated on a review of the team reflexivity literature and the definitions of relevant team reflective behaviors.

To begin with, we decided to not only code the number of utterances with a reflection-based content, but also their length (i.e., number of letters coded as reflection). We chose this procedure in order to better capture the quantity (and hence, quality) of reflection. For instance, simply accounting for the number of reflective utterances would mean to weight a long and a short sentence equally, possibly underestimating the effects of long utterances. Therefore, considering the length of an utterance allows to better capture quantity (and in return quality) of reflection.

Furthermore, we decided to conceive team reflection as a process focused on evaluating activities (cf. Schippers et al., 2007; Gabelica, Van den Bossche et al., 2014; Otte et al., 2017). It can therefore be defined as “aggregating pieces of information and making interpretations that go beyond their individual significance in order to draw more elaborate conclusions and develop new or different methods and strategies” (Otte et al., 2017, p. 301). Hence, based on the coding scheme proposed in Gabelica, Van den Bossche et al. (2014), and the items from the information evaluation dimensions of the REMINT, we constructed the coding scheme for quality and quantity of team reflection as the following. First, we conceptually separated quantity and quality of team reflection. We conceived quantity of team reflection as the total amount of reflection behaviors coded in each group. We then delineated quality of team reflection into three different levels (i.e., categories) and conceived qualitative team reflection as ratio of level three (i.e., deep reflection) and level one (i.e., superficial reflection) team reflection (cf. West, 2000).

Level one included global and superficial evaluations, including global performance ratings (e.g., “I think we performed quite well”), or ratings of the strategy the team chose (e.g., “I liked our strategy”). Level two included evaluations of subcomponents of overall team performance, as for example strategies (e.g., “Using the new way of calculating the values for each city was a good idea”). Conclusively, level 3 contained (1) descriptions and evaluations of particular behaviors or processes, including why and in which way these processes affected overall team performance (e.g., “I liked that you distributed the tasks within the team, because that really helped us to coordinate ourselves more effectively”), and (2) evaluations of particular processes that negatively affected team performance (e.g., “I have difficulties coordinating all the tasks I have to manage”).

All team reflection sessions where coded twice by three raters. In case of diverging assessments, the respective case was discussed until agreement was reached.

3.6.5 Data Analysis Strategy

We used structural equation modeling for testing our hypotheses. Hypotheses 1a and 1b (Figure 3.1), as well as Hypotheses 2 to 7 (see Figures 3.2 and 3.3) were tested in two separate models. For testing Hypothesis 2 to 7, we integrated an auto-regressive path between the first round and second round team performance. Autoregressive effects describe the effect of a construct on itself over time (Selig & Little, 2012), and are often referred to as the stability of the construct. A large auto-regressive coefficient indicates very little change over time, whereas a small coefficient indicates substantial change. Consequently, in our study, the auto-regressive path can be interpreted as controlling for previous team performance, and therefore whether the change in team performance can be related to the effects of team reflection. This approach allows ruling out that the relation between team reflection and team performance is solely based on the fact that better teams always show generally higher levels of reflection. Finally, since we expected a substantial correlation between quality and quantity

of team reflection, we allowed the error terms for both values to covariate in order to improve model fit. Furthermore, following the guidelines outlined in Cole et al. (2011) for highly correlated variables, we controlled whether the interaction term of quality and quantity of team reflection impacted path estimates when both constructs were simultaneously related to a criterion.

Furthermore, the sample size on the team level (N = 46) may produce biased estimates when using maximum likelihood estimations (cf. Hinkin, 1998). We therefore used Bayesian structural equation modeling (BSEM) (see Muthén & Asparouhov, 2012; Zyphur & Oswald, 2015). BSEM has several advantages over common maximum likelihood estimations, including that it allows for unbiased estimates in smaller sample sizes (Song & Lee, 2012). We employed two Markov chain Monte Carlo chains and set the number of iterations to 100,000 to ensure model stability. To assess model fit, we used the posterior predictive p-value (PPP) and the posterior predictive checking (PPC) 95% credibility interval (CI), which indicates whether the data fits the proposed model. A PPP value above .05 indicates an acceptable fit, whereas a PPP value of .5 indicates an excellent fit (Muthén & Asparouhov, 2012). The PPC CI should include a negative lower limit and a positive upper limit (Zyphur & Oswald, 2015) to indicate a good fit. We used the BIC (Bayesian information criterion) and the DIC (Deviance information criterion) for model comparisons (cf. Kaplan & Depaoli, 2012). A benefit of using the BIC for model comparison is that models do not need to be nested (Kaplan & Depaoli, 2012). In general, a lower BIC and DIC can be interpreted as model that has more likely to have generated the data (West, Taylor, & Wu, 2012). Finally, we examined whether the solution was stable when we increased the number of iterations. The statistical analysis was conducted in Mplus 7.4 (Muthén & Muthén, 2017), whereas the transcription and coding of the data was conducted in MAXQDA (VERBI Software, 2017).

3.7 Results

3.7.1 Preliminary Analysis

Means, standard deviations, reliabilities alpha, multi-item agreement indices $r^*_{WG(J)}$ (Lindell, Brandt, & Whitney, 1999) and $a_{WG(J)}$ (Brown & Hohenstein, 2005), as well as the correlations among the study variables are depicted in Table 3.1.

The team implementation agreement indices ($r^*_{WG(J)}$ and $a_{WG(J)}$) were rated as very strong agreement (LeBreton & Senter, 2008), however, the team reflection agreement values can only be interpreted as moderate agreement. Moreover, although the $a_{WG(J)}$ is conceived as a more conservative estimate of agreement (Brown & Hohenstein, 2005), the values obtained with this index were higher than the values obtained with $r^*_{WG(J).}$ However, when the number of judges and the number of items is fairly small, which was the case in our study, $r^*_{WG(J)}$ may be attenuated, in particular in combination with lower reliability values (LeBreton & Senter, 2008). Yet, LeBreton and Senter argued that aggregating all groups can be done, as long some of the groups have sufficiently high values. Inspections of all $r^*_{WG(J)}$ values revealed that 21 (46%) teams on the quantitative team reflection dimension, and 15 (33%) teams on the qualitative team reflection dimension had sufficiently high agreement values ($r^*_{WG(J)} \geq 0.7$). Moreover, agreement values of the teams were not correlated to team performance, or quality and quantity of team reflection. Hence, we continued our analysis with the full sample of 46 teams.

To examine discriminant validity, we conducted a team-level Bayesian confirmatory factor analysis including the team reflection quality/quantity, as well as implementation measure. Following the procedures outlined in Otte et al. (2017), we allowed residual terms of the respective qualitative and quantitative items of the team reflection scales to covary. Resulting model-fit indices were good (PPP = 0.28, PPP 95% CI = [-24.83 | 43.60]).

Table 3.1: Correlations Among and Descriptive Statistics for Questionnaire-based Data

		M (*SD*)	$r^*_{WG(j)}$	$a_{WG(J)}$	1	2	3	4	5	6	7
1	Performance T_0	1.09 (1.29)	--	--	--						
2	Self-rated Reflection Quantity. t_1	2.70 (0.55)	.58	.68	-.33*	**.81**					
3	Self-rated Reflection Quality. t_1	2.58 (0.48)	.52	.59	-.26*	.75**	**.77**				
4	Other-rated Reflection Quantity t_1	378.30 (32.12)	--	--	.11	-.15	-.08	--			
5	Other-rated Reflection Quality t_1	0.96 (1.63)	--	--	.24	-.28*	-.02	.63**	--		
6	Implementation t_1	4.65 (0.42)	.89	.97	.02	-.17	-.07	.12	.20	**.95**	
7	Performance t_2	4.71 (2.52)	--		.39**	-.47**	-.17	.20	.42**	.43**	--

Notes. N_{Team} = 46. Reliability alphas on the diagonal (bold).

*p < .05. **p < .01 (one-tailed).

As expected, factor correlations between the qualitative and quantitative dimensions were high (β = 0.84, posterior S.D. = 0.08, 95% CI [0.64 | 0.95], $p < .01$) but below the criterion of 0.85 (cf. Cohen, Cohen, West, & Aiken, 2003).

Team performance increased from t_0 to t_1 (see also Table 3.1). The performance value calculated from number of correctly computed value-points and correctly assigned clients, increased from 1.08 in the first, to 4.71 in the second round. In average, participants correctly assigned 0.56 clients in the first, and 1.3 clients in the second round. Regarding the coding of the reflection data, we overall coded 148 utterances as level 1 team reflection, 60 as level 2 team reflection, and 34 as level 3 team reflection. Correlations between the number of letters coded for each level and performance are displayed in Table 3.2.

Table 3.2: Audio-data based correlations of levels of team reflection and performance

		M (*SD*)	1	2	3	4	5
1	Performance t_0	1.09 (1.29)	--				
2	Reflection t_1: Level 1	139.33 (98.77)	-.03	--			
3	Reflection t_1: Level 2	118.91 (139.55)	.16	.20	--		
4	Reflection t_1: Level 3	120.07 (186.82)	.08	.24	.47**	--	
5	Performance t_2	4.71 (2.52)	.39**	-.06	.09	.30*	--

Notes. N_{Team} = 46. Perf. = Performance. Level 1 = Number of letters coded as level 1 reflection. Level 2 = Number of letters coded as level 2 reflection. Level 3 = Number of letters coded as level 3 reflection.

*$p < .05$. **$p < .01$ (one-tailed).

3.7.2 Hypothesis Testing

Hypothesis 1a predicted a moderate correlation between self-rated and other-rated quantitative reflection. This hypothesis was not confirmed (see Figure 3.1), as there was no significant correlation between self-rated and other-rated quantity of team reflection.

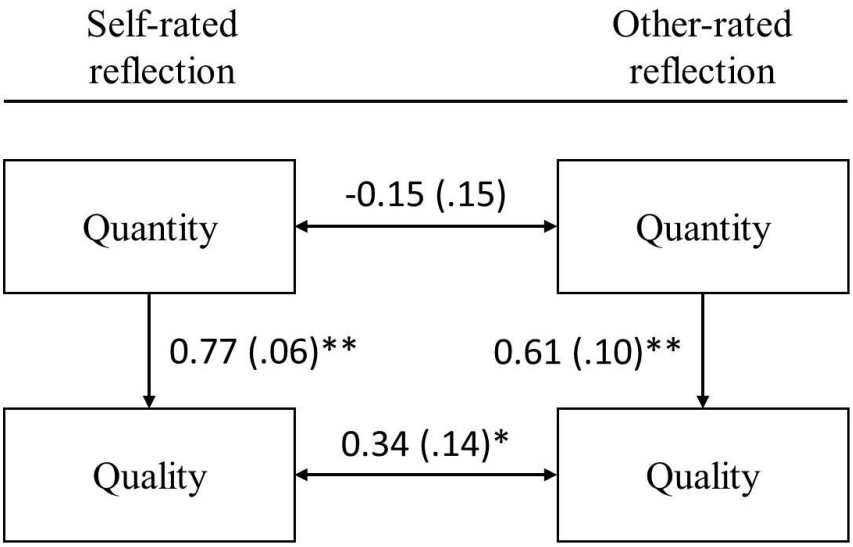

Figure 3.1. Relation between self-rated and other-rated team reflection.

Note. N = 46. PPP = 0.41. PPP 95% CI = [-13.79 | 17.09]. All estimates are standardized. Self- and other-rated qualitative team reflection are controlled for the respective quantitative team reflection measure, as indicated by the unidirectional arrows.

*p < .05. **p < .01 (one-tailed).

Hypothesis 1b predicted a moderate relation between self-rated and other-rated qualitative team reflection, controlled for quantitative team reflection. This hypothesis was confirmed (β = 0.34, posterior S.D. = 0.14, 95% CI [0.04 | 0.59], p < 0.05). The following hypotheses were based on the structural equation models depicted in Figure 3.2, which is based on self-rated team reflection data, and Figure 3.3, which is based on other-rated team reflection data. Overall, model fit indices for both models were excellent, with a PPP of 0.49 and confidence interval that included zero (PPP_{Self} 95% CI = [-17.56 | 18.40]; PPP_{Other} 95% CI = [-17.28 | 18.15]). The following reported path estimates are standardized values, and corresponding p-values are one-tailed, according to our hypotheses.

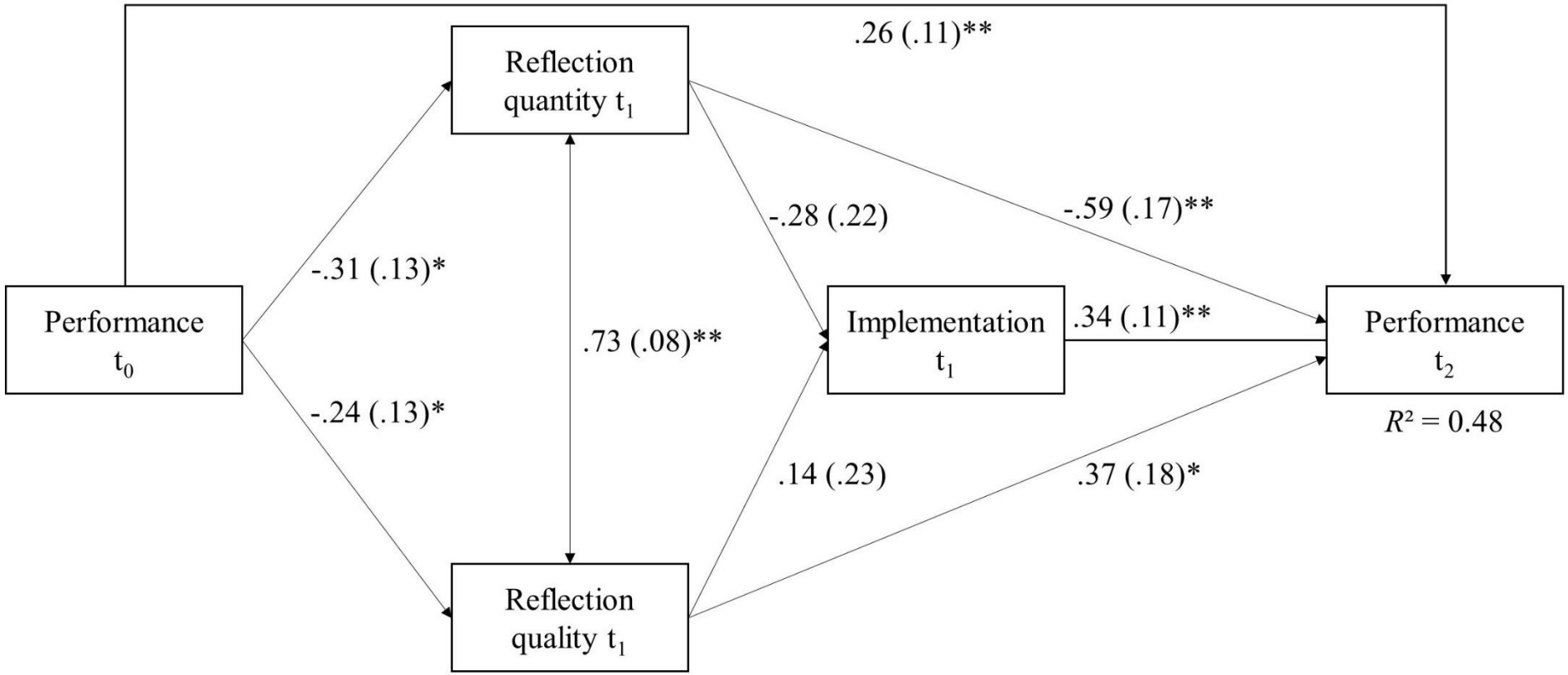

Figure 2: Model using self-rated team reflection measures.

Note. $N = 46$. All estimates are standardized.

*$p < .05$. **$p < .01$ (one-tailed).

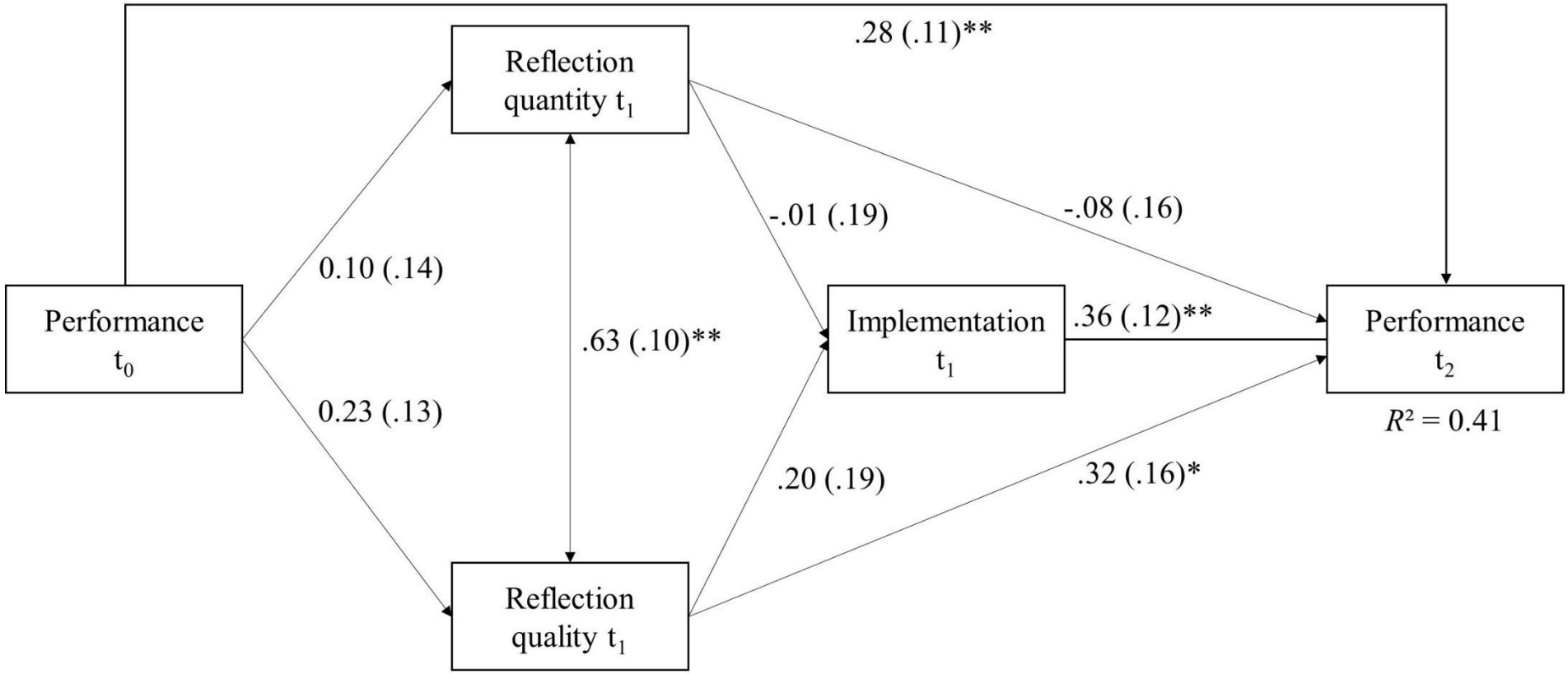

Figure 3: Model using other-rated team reflection measures.

Note. $N = 46$. All estimates are standardized.

*$p < .05$. **$p < .01$ (one-tailed).

Hypothesis 2 predicted that team performance is negatively related to subsequent qualitative and quantitative team reflection (self- and other-rated). The hypothesis was partly confirmed. For the self-rated data, we found a significant negative relation between first round team performance (t_0) and subsequent quantitative ($\beta = -0.31$, posterior S.D. = 0.13, 95% CI [-0.53 | -0.03], $p < 0.05$) and qualitative ($\beta = -0.24$, posterior S.D. = 0.13, 95% CI [-0.48 | 0.040], $p < 0.05$) team reflection. For the other-rated data, we did not find any significant relations at team performance at t_0 and subsequent reflection.

Hypothesis 3 predicted a negative relation between quantitative team reflection and performance improvement (controlled for qualitative reflection). The hypothesis was also partly confirmed. Self-rated quantitative team reflection was strongly negatively related to team performance improvement ($\beta = -0.59$, posterior S.D. = 0.17, 95% CI [-0.92 | -0.23], $p < 0.001$). We found no significant relations for the other-rated data.

Hypothesis 4 predicted a positive relation between qualitative team reflection and performance (controlled for quantitative reflection). The hypothesis was confirmed. Qualitative team reflection was positively related to subsequent team performance improvement for both self-rated ($\beta = 0.37$, posterior S.D. = 0.18, 95% CI [0.02 | 0.72], $p < 0.05$) and other-rated data ($\beta = 0.32$, posterior S.D. = 0.16, 95% CI [-0.00 | 0.62], $p < 0.05$).

Hypothesis 5 predicted a partial mediation of implementation between team reflection and performance improvement. This hypothesis was not confirmed. Although we did find significant relations between implementation and performance improvement that explained unique variance above quality and quantity of team reflection in the self-rated team reflection model ($\beta = 0.37$, posterior S.D. = 0.18, 95% CI [0.02 | 0.72], $p < 0.05$) and the other-rated team reflection model ($\beta = 0.36$, posterior S.D. = 0.12, 95% CI [0.11 | 0.58], $p < 0.01$), we did not find any significant relation between either quality nor quantity of reflection and implementation. Hence, although the estimates in the model based on self-rated team

reflection pointed into the right directions, we also couldn't confirm Hypothesis 6 and 7, which predicted a negative relation between quantitative reflection and subsequent implementation, and a positive relation between qualitative reflection and subsequent implementation.

Finally, we proposed in our research question to examine whether other-rated team reflection can explain performance above self-rated team reflection. However, a regression-based analysis replicating the relevant parts from the structural equation model showed that this was not the case (see Table 3.3).

Table 3.3: Incremental validity of other-rated team reflection

	Performance t_2			
Variable	Model 1	Model 2	Model 3	Model 4
Performance t_0	.38*	.25*	.22*	.24*
Self-rated reflection quantity t_1		-.72**	-.60**	-.54**
Self-rated reflection quality t_1		.45*	.35*	.32*
Other-rated reflection quantity t_1			-.03	-.03
Other-rated reflection quality t_1			.23	.18
Implementation t_1				.32**
Adjusted R^2	.12	.32	.32	.42
R^2	.14	.37	.32	.50
Change R^2	.14*	.23**	.04	.10**

Notes. $N_{Team} = 46$.

*$p < .05$. **$p < .01$ (one-tailed).

3.8 Discussion

The purpose of the current research was to examine the relations between team performance, subsequent self- and other-rated quality and quantity of team reflection, implementation, and successive performance. Overall, we found considerable support for our hypotheses. However, three key-findings are particularly relevant.

First, when time is scarce, qualitative team reflection is positively, whereas quantitative team reflection is negatively related to team performance improvement. Hence, this study has shown for the first time that measuring only the quantity of team reflection leads to the inconclusive results observed in previous research (e.g., De Dreu, 2002; Wiedow & Konradt, 2011). Thus, it is crucial to take both team reflection dimensions simultaneously into account in order to disentangle the relationship between team reflection and performance.

The second key finding is that we were only able to reproduce the qualitative relation to team performance, when using other-rated team reflection data. However, other-rated quantitative team reflection was neither significantly correlated to the self-rated quantity measure (see Figure 3.1), nor did not show a negative relation to team performance (see Figure 3.3). This finding is particularly surprising, as one would expect that assessing the quantity of team reflection would be easier for both participants and investigators. As this is obviously not the case, future research should explore reasons for the missing relation.

An initial explanation might be that the questionnaire in this study did not set a frame of reference for the term ‘often’ that was used when measuring quantity via self-ratings. Teams might, however, have different perceptions regarding the amount of reflection. This assumption is supported by the relatively low agreement values we obtained, indicating substantial differences in the perceptions of the team members. Hence, a possible solution for future research in laboratory settings is to set a frame for the term ‘often’ (e.g., “regarding the last five minutes, how many of these minutes did you spend reflecting”).

The second factor that might have contributed to the missing relation could have been the ratio of reflection taking place in discussions (i.e., open or visible reflection), and reflection taking place within each team member (i.e., invisible reflection). Inspection of audio data revealed that several teams that remained totally silent for short periods of time. Stemming our analysis on linguistic utterances, we were thus not able to decide whether the

team members reflected individually during this time, or whether they were busy with for example developing plans for the next round. It is therefore possible that we did not capture the individual element in team reflection, which limits the validity of the behavioral data.

Thirdly, we only transcribed the designated team reflection phases. However, teams might have already reflected during the action phases. Consequently, the reflection we observed might have only been an extract from the team reflection that already started during the action phase, leading to an underestimation of the quantity of reflection.

The third key-finding of this study is the missing relation between team reflection and implementation. Yet, although both self-rated and other-rated team reflection was not related to implementation, the estimates pointed into the expected directions. This finding may therefore partly be attributed to the rather small sample size, in combination with comparatively high standard deviations for the respective relations (see Figures 3.2 and 3.3). Hence, we might have lacked the statistical power to detect the expected effects. In addition, the mean implementation (i.e., living up to agreements) value was fairly high, with about 40% the teams reporting the highest possible implementation value. As the reported team reflection values were considerably lower, substantial amounts of variance in implementation might have been lost, thereby further underestimating the relationship between team reflection and implementation.

Additionally, measuring implementation as living up to agreements does not take the quality of the developed action-plan into account. Hence, teams might produce various unpractical solutions during the reflection (and planning) sessions that are discarded, or even fail to be implemented during the action-process, resulting in additional variance in implementation that is not explained by preceding team reflection.

Finally, the conceptualization of team reflection as inevitable step prior to implementation might be too naïve. For instance, Gabelica, Van den Bossche et al. (2014)

reported that teams seldom engaged in full cycles of reflexive behaviors (i.e., reflecting, planning, and implementing). Teams may therefore implement changes without prior reflection (or even planning). In fact, West (2000) referred to non-reflexive teams as teams that act spontaneously without awareness of the action. Hence, it might be a more rational step to defer the normative theoretical assumptions regarding what teams *should do*, and instead rather examine what they *actually do*. Hence, in spite of the promising results that we obtained with questionnaires, a stronger emphasize on behaviorally oriented research drawing on audio or video data could provide the necessary insights that are necessary to develop more detailed theories, and more useful practical recommendations.

3.8.1 Managerial Implications

The results of this study also have managerial implications. At first, it is worth mentioning that managers should not only encourage teams to reflect, but also guide them to focus on the appropriate reflexive behaviors (i.e., qualitative or quantitative focus). Moreover, the findings of this research indicate that when time is scarce, the focus should be put upon qualitative reflection, as this approach would allow teams utilize their available resources (i.e., time) more effectively.

Moreover, as teams sometimes tend to focus on task irrelevant contents in reflection sessions (Gurtner et al., 2007), guidance and facilitation (e.g., guided team reflection, Konradt et al., 2015) of team reflection through leadership (cf. West, 2000), or coaching (cf. Hackman & Wageman, 2005), may require a guide or facilitator to draw conclusions from observations during the team reflection sessions. However, in order to give useful advice, team reflection facilitators may need trainings that enable them to not only guide a team towards team reflection, but also to aid them to distinguish between superficial and deep team reflection. The three levels of depth of team reflection that we proposed in this study might therefore give useful advice.

3.8.2 Limitations and Future Research

Despite the rigorous study design and encouraging findings, this study holds five limitations that restrict the findings, which future research may endeavor to address.

The first limitation concerns the relation between quantity of team reflection and team performance. Although we did find a strong negative effect of quantitative team reflection on performance, we cannot rule out that quantitative reflection might only be the symptom, and not the cause of low performance. Although this scenario is unlikely in the university context, teams might have lacked the mental ability to deeply evaluate their performance, because they did not understand the task. Future research should therefore also explore the conditions that facilitate or hinder team reflection, as for instance general mental ability.

The second limitation arises from the fairly low agreement values that we obtained from the self-rated team reflection measures. Although 30-40% of all teams showed sufficiently high agreement, multiple teams exhibited extremely low agreement regarding the question how much and how deep they reflected. It is assumed in multi-level research that the variance of group member's responses can provide additional value above aggregated mean judgements (Cole et al., 2011). Hence, future research should explore potential differences between the referent-shift consensus measures (i.e., "we") used in this study, and direct consensus (i.e., "I") measures that might explain additional variance (Chan, 1998).

Thirdly, in spite of the rigorous development process of our team reflection coding scheme, we cannot completely rule out that the missing relation between self- and other-rated quantitative team reflections could also be attributed to miss-conceptualizations of other-rated team reflection. However, as the other-rated qualitative team reflection does correlate with the self-rated measure, and is similarly related to team performance, we'd rather argue in favor of our coding scheme.

As a fourth limitation, we did only examine the influence of a temporal contingency on quality and quantity of team reflection. However, other contingencies might shift the effects of qualitative and quantitative reflection. For instance, giving teams more time to reflect above a more extensive task that provides more possibilities to incrementally optimize strategies could lead to positive effects of quantitative reflection, as is would allow teams to cover more different subjects. In contrast, qualitative reflection might in this particular scenario even hinder teams to cover all relevant elements. Future research should therefore explore the effects of qualitative and quantitative team reflection in other scenarios that are based on other principles and rules.

Finally, as previous studies based on the weather company task (Wiedow et al., 2013; Konradt et al., 2015), this study was also used a laboratory setting with a sample of relatively highly educated students that mainly studied at the psychology department. Thus, this might limit the external validity of our results (cf. Diopboye, 1990). Future research should examine whether the results can be generalized to the organizational field.

3.9 References

Berelson, B. (1952). *Content Analysis in Communication Research*. Glencoe, IL: The Free Press.

Berry, C. M., Carpenter, N. C., & Barratt, C. L. (2012). Do other-reports of counterproductive work behavior provide an incremental contribution over self-reports? A meta-analytic comparison. *Journal of Applied Psychology, 97*, 613-636. doi:10.1037/a0026739

Brown, R. D., & Hauenstein, N. M. A. (2005). Interrater agreement reconsidered: An alternative to the rwg indices. *Organizational Research Methods, 8*, 165–184. doi:10.1177/1094428105275376

Carpenter, N. C., Berry, C. M., & Houston, L. (2014). A meta-analytic comparison of self-reported and other-reported organizational citizenship behavior. *Journal of Organizational Behavior, 35*, 547-574. doi:10.1002/job.1909

Carver, C.S., & Scheier, M.F. (1998). *On the Self-Regulation of Behavior*. Cambridge, England: Cambridge University Press.

Chan, D. (1998). Functional relations among constructs in the same content domain at different levels of analysis: A typology of composition models. *Journal of Applied Psychology, 83*, 234-246. doi:10.1037/0021-9010.83.2.234

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

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (3rd ed.). Mahway, NJ: Lawrence Erlbaum Associates.

Cole, M. S., Bedeian, A. G., Hirschfeld, R.R., & Vogel, B. (2011). Dispersion-composition models in multilevel research: A data-analytic framework. *Organizational Research Methods, 14*, 718-734. doi:10.1177/109442811038907

De Dreu, C. K. W. (2002). Team innovation and effectiveness: The importance of minority dissent and reflexivity. *European Journal of Work and Organizational Psychology, 11*, 285–298. doi:10.1080/13594320244000175

DeShon, R.P., Kozlowski, S.W. J., Schmidt, A.M., Milner, K.R., & Wiechmann, D. (2004). A multiple goal, multilevel model of feedback effects on the regulation of individual and team performance. *Journal of Applied Psychology, 89*, 1035–1055. doi:10.1037/0021-9010.89.6.1035

Dipboye, R. L. (1990). Laboratory vs. field research in industrial organizational psychology. In C. L. Cooper & I. T. Robertson (Eds.), *International Review of Industrial and Organizational Psychology* (Vol. 5, pp. 1–34). New York, NY: Wiley.

Förster, J., Higgins, E. T., & Bianco, A. T. (2003). Speed/accuracy decisions in task performance: Built-in trade-off or separate strategic concerns? *Organizational Behavior and Human Decision Processes, 90*, 148-164. doi:10.1016/S0749-5978(02)00509-5

Gabelica, C., Van den Bossche, P., De Maeyer, S., Segers, M., & Gijselaers, W. (2014). The effect of team feedback and guided reflexivity on team performance change. *Learning and Instruction, 34*, 86–96. doi:10.1016/j.learninstruc.2014.09.001

Gabelica, C., Van den Bossche, P., Segers, M., & Gijselaers, W. (2014). Dynamics of team reflexivity after feedback. *Frontline Learning Research, 2*, 64–91. doi:10.14786/flr.v2i2.79

Gurtner, A., Tschan, F., Semmer, N. K., & Nägele, C. (2007). Getting groups to develop good strategies: Effects of reflexivity interventions on team process, team performance, and shared mental models. *Organizational Behavior and Human Decision Processes, 102*, 127-142. doi:10.1016/j.obhdp.2006.05.002

Hackman, J. R., & Wageman, R. (2005). A theory of team coaching. *Academy of*

Management Review, 30, 269-287. doi:10.5465/AMR.2005.16387885

Heidemeier, H. & Moser, K. (2009). Self-other agreement in job performance ratings: A meta-analytical test of a process model. *Journal of Applied Psychology, 94*, 353-370. doi:10.1037/0021-9010.94.2.353

Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods, 1*, 104-121. doi:10.1177/109442819800100106

Ilgen, D. R., Hollenbeck, J. R., Johnson, M., & Jundt, D. (2005). Teams in organizations: from input-process-output models to IMOI models. *Annual Review of Psychology, 56*, 517–543. doi: 10.1146/annurev. psych.56.091103.070250

Kaplan, D., & Depaoli, S. (2012). Bayesian structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of Structural Equation Modeling* (pp. 650-673). New York: The Guildford Press.

Konradt, U., & Eckardt, G. (2016). Short-term and long-term relations among reflection and performance in teams: Evidence from a four-wave longitudinal study. *European Journal of Work and Organizational Psychology, 25, 804–818.* doi:10.1080/1359432X.2016.1160058

Konradt, U., Otte, K.-P., Schippers, M. C., & Steenfatt, C. (2016). Reflexivity in teams – A review and new perspectives. *The Journal of Psychology: Interdisciplinary and Applied, 150*, 153–174. doi:10.1080/00223980.2015.1050977

Konradt, U., Schippers, M. C., Garbers, Y., & Steenfatt, C. (2015). Effects of guided reflexivity and team feedback on team performance improvement: The role of team regulatory processes and cognitive emergent states. *European Journal of Work and Organizational Psychology, 24*, 777–795. doi:10.1080/1359432X.2015.1005608

LeBreton, J. M., & Senter, J. L. (2007). Answers to 20 questions about interrater reliability

and interrater agreement. *Organizational Research Methods*, *11*, 815-852. doi:10.1177/1094428106296642

Lindell, M. K., Brandt, C. J., & Whitney, D. J. (1999). A revised index of agreement for multi-item ratings of a single target. *Applied Psychological Measurement, 23*, 127-135.

Lu, L., Yuan, Y. C., & McLeod, P. L. (2012). Twenty-five years of hidden profiles in group decision making: A meta-analysis. *Personality and Social Psychology Review, 16*, 54-75. doi:10.1177/1088868311417243

Marks, M. A., Mathieu, J. E., & Zaccaro, S. J. (2001). A temporally based framework and taxonomy of team processes. *Academy of Management Review, 26*, 356–376. doi:10.5465/AMR.2001.4845785

Moreland, R. L., & McMinn, J. G. (2010). Group reflexivity and performance. *Advances in Group Processes, 27*, 63–95. doi:10.1108/s0882-6145(2010)0000027006

Muthén, B. & Asparouhov, T. (2012). Bayesian SEM: A more flexible representation of substantive theory. *Psychological Methods, 17*, 313-335. doi:10.1037/a0026802

Muthén, L. K., & Muthén, B. O. (2016). *Mplus User's Guide. Seventh Edition*. Los Angeles, CA: Muthén & Muthén.

Nadler, D.A. (1979). The effects of feedback on task group behavior: A review of the experimental research. *Organizational Behavior and Human Performance, 23*, 309-338. doi:10.1177/1059601194193006

Norman, D. A., & Bobrow, D. G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology, 7*(1), 44-64.

Otte, K.-P., Konradt, U., Garbers, Y., & Schippers, M. C. (2017). Development and validation of the REMINT: a reflection measure for individuals and teams. *European Journal of Work and Organizational Psychology, 26,* 299-313.

doi:10.1080/1359432X.2016.1261826

Paulhus, D. L., & Vazire, S. (2007). The self-report method. In R. W. Robins, R. C. Fraley, & R. F. Krieger (Eds.), *Handbook of Research Methods in Personality Psychology* (pp. 224-239). New York: The Guilford Press.

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88,* 879-903. doi:10.1037/0021-9010.88.5.879

Podsakoff, P. M., & Organ, D. W. (1986). Self-reports in organizational research: Problems and prospects. *Journal of Management, 12,* 531–544. doi:10.1177/014920638601200408

Rourke, L., & Anderson, T. (2004). Validity in quantitative content analysis. *Education Technology Research and Development, 52*, 5-18. doi:10.1007/BF02504769

Schippers, M. C., Den Hartog, D. N., & Koopman, P. L. (2007). Reflexivity in teams: A measure and correlates. *Applied Psychology: An International Review, 56,* 189–211. doi:10.1111/j.1464-0597.2006.00250.x

Schippers, M. C., Edmondson, A. C., & West, M. A. (2014). Team reflexivity as an antidote to team information-processing failures. *Small Group Research, 56,* 731–769. doi:10.1177/1046496414553473

Schippers, M. C., Homan, A. C., & van Knippenberg, D. (2013). To reflect or not to reflect: Prior team performance as a boundary condition of the effects of reflexivity on learning and final team performance. *Journal of Organizational Behavior, 34,* 6–23. doi:10.1002/job.1784

Schippers, M. C., West, M. A., & Dawson, J. F. (2015). Team reflexivity and innovation: The moderating role of team context. *Journal of Management, 41,* 769–788.

doi:10.1177/0149206312441210

Selig, J. P., & Little, T. D. (2012). Autoregressive and cross-lagged panel analysis for longitudinal data. In B. Lausen, T. D. Little, & N. A. Card, *Handbook of Developmental Research Methods* (pp. 265-278). New York: The Guildford Press.

Smith-Crowe, K., Burke, M. J., Cohen, A., & Doveh, E. (2014). Statistical significance criteria for the rWG and average deviation interrater agreement indices. *Journal of Applied Psychology, 99*, 239–261. doi: 10.1037/a9934556.

Song, X. Y., & Lee, S. Y. (2012). *Basic and Advanced Bayesian Structural Equation Modeling: With Applications in the Medical and Behavioral Sciences*. London: Wiley.

Stasser, G., & Titus, W. (1985). Pooling of unshared information in group decision making: Biased information sampling during discussion. *Journal of Personality and Social Psychology, 48,* 1467-1478. doi:10.1037/0022-3514.48.6.1467

VERBI Software (2017). MAXQDA Base (Version 12) Software for Qualitative, Quantitative and Mixed Methods Research. Berlin: VERBI GmbH.

West, M. A. (2000). Reflexivity, revolution, and innovation in work teams. In M. M. Beyerlein, D. Johnson, & S. T. Beyerlein (Eds.), *Product Development Teams* (Vol. 5, pp. 1–29). Stamford, CT: JAI Press

West, S. G., Taylor, A. B., & Wu, W. (2012). Model fit and model selection in structural equation modelling. In R. H. Hoyle (Ed.), *Handbook of Structural Equation Modelling* (pp.). New York: Guilford.

Wiedow, A., & Konradt, U. (2011). Two-dimensional structure of team process improvement: Team reflection and team adaptation. *Small Group Research, 42,* 32–54. doi:10.1177/1046496410377358

Wiedow, A., Konradt, U., Ellwart, T., & Steenfatt, C. (2013). Direct and indirect effects of team learning on team outcomes: A multiple mediator analysis. *Group Dynamics:*

Theory, Research, and Practice, *17*, 232-251. doi:10.1037/a0034149

Zyphur, M. J. & Oswald, F. L. (2015). Bayesian estimation and inference: A user's guide. *Journal of Management, 41,* 390–420. doi:10.1177/0149206313501200

Kapitel 4: Studie 3

Revisiting punctuated equilibrium:

a longitudinal examination of the dynamic relation of team reflection and performance

Otte, K.-P., Konradt, U., & Eckardt, G. (2017). *Revisiting punctuated equilibrium: a longitudinal examination of the dynamic relation of team reflection and performance*. Unpublished manuscript.

Abstract

Punctuated equilibrium theory (Gersick, 1991; Tushman & Romanelli, 1985) posits that development and change in organizations can be conceived as long periods of stability (equilibrium), that are punctuated by sudden and disruptive periods of revolution and upheaval. Organizational research has, however, rarely explored the process mechanisms that support or hinder adaptation in punctuated equilibrium. To overcome this shortcoming, we refine the punctuated equilibrium theory by including recent episodic team effectiveness frameworks, and input-mediator-output-input models. Moreover, we argue that the effects of team processes on outcomes are contingent on the changing demands that result from revolutionary and equilibrium periods. We hypothesize that team processes associated with exploration strategies are beneficial for performance during revolutionary, and detrimental during equilibrium periods. On the contrary, we propose that exploitation strategies are beneficial during equilibrium, and detrimental during revolutionary periods. To test these assumptions, we leverage upon the process of team reflection and its qualitative (exploration-focused behavior) and quantitative (exploitation-focused behavior) subdimensions. Using Bayesian structural equation modeling, we show in a six-wave longitudinal study with 86 teams working on a complex business simulation task, that the effects of quality and quantity of team reflection are contingent on the demands that arise from revolutionary and equilibrium periods. The results partly support our hypotheses and additions to punctuated equilibrium theory. Moreover, they provide relevant insights into the team reflection-performance relationship as important mechanism in team adaptation. Theoretical and practical implications are discussed.

4.1 Introduction

At the beginning of the 21st century, several electronic product categories (e.g., mobile phones, mp3 players, digital cameras, or PDAs) were largely dominated by relatively few organizations that produced small and incremental upgrades of previous technologies. This fairly stable process of incremental upgrades was heavily disrupted by the introduction of smartphones, which replaced multiple different products and integrated their functions within one single device. While some organizations could adapt to this change, others lost not only their industry-leading positions, but also billions of dollars. In recent times, smartphone-technology arrived at a state that is again not only dominated by incremental upgrades, but also by very few organizations.

Tushman and Romanelli (1985) and Gersick (1991), among others, referred to this type of development as *punctuated equilibrium*. Based on assumptions from evolutionary biology, the punctuated equilibrium theory posits that organizations exist in an environment shaped by long periods of stability (equilibrium), and relatively brief periods of revolution and upheaval (punctuations). However, despite the plausibility of the concept, studies investigating its validity (e.g., Okhuysen & Waller, 2002) focused almost exclusively on the question whether team develop in experimental settings according to the assumptions from punctuated equilibrium. Organizational research has, to our knowledge, rarely examined how punctuated equilibrium unfolds over a longer time period, as it was originally proposed by for instance Tushman and Romanelli (1985).

Moreover, a crucial shortcoming of the theory itself, is its focus on the structural component of change. The structure is thereby conceived as reoccurring patterns of punctuations and equilibria, that require teams to develop and adapt (i.e., altering behaviors to meet environmental demands, Pulakos, Arad, Donovan, & Plamondon, 2000). Underlying process mechanisms, however, that determine how effectively teams develop and adapt in

periods of revolution or equilibrium, have been left largely untouched by previous theory and research. Consequently, although the theory provides an advanced perspective on change, it offers surprisingly little insights into the dynamics of longitudinal developments of teams within the theoretical frame it provides (Humphrey & Aime, 2014).

In order to overcome these shortcomings, we propose to extend Gersick's (1991) and Tushman and Romanelli's (1985) line of work and expand the theory through the following additions. First, as teams have become the building blocks of modern organizational designs (Mathieu, Tannenbaum, Donsbach, & Alliger, 2014), we conceive teams as the focal unit of analysis within the punctuated equilibrium theory. Secondly, to provide a theoretical foundation for longitudinal change of processes in teams, we integrate the multiphase episodic framework provided by Marks, Mathieu, and Zaccaro (2001), and input-mediator-output-input models (IMOI; Ilgen, Hollenbeck, Johnson, & Jundt, 2005). Thirdly, we argue that the reciprocal relationships between team processes and outcomes among episodes depend on the varying environmental demands that emerge from theory's conception of change. To better capture these dynamic features, we propose a contingency perspective (Lawrence & Lorsch, 1967) on the relations between processes and outcomes in punctuated equilibrium. In this perspective, it is assumed that team processes do not always affect outcomes in a similar manner, but are contingent on other variables (e.g., environmental demands).

In the following, we provide further elaboration of the suggested additions. Furthermore, we empirically examine the processes of qualitative and quantitative team reflection (Otte, Konradt, Garbers, & Schippers, 2017) and how they influence the success of team adaptation in an environment shaped by punctuated equilibrium.

4.2 The Punctuated Equilibrium Theory

Models and theories for understanding the development of teams are generally listed in two categories: life-cycle approaches and equilibrium approaches (Humphrey & Aime, 2014). Lifecycle approaches, as for example Tuckman's five phases of team development (forming, storming, norming, performing, and adjourning; see Tuckman & Jensen, 1977) mainly assume that teams evolve through the same linear cycles of starting-up, growing, harvesting, terminating, and starting up again (Bartnunek & Woodman, 2015; Weick & Quinn, 1999). Thus, they are not particularly flexible and cannot consider external influences (i.e., changing environments) that requires teams to adapt (Arrow, 1997).

In a different perspective, teams are therefore not assumed to develop through reoccurring life-cycles, but as self-regulating entities that develop towards relatively stable configurations. Extending upon this idea, Tushman and Romanelli (1985), and Gersick (1991) developed the punctuated equilibrium theory for describing team (and organizational) development in dynamic environments. However, in this research we will focus on teams as focal entity.

Within the theory, the authors distinguish between (1) *equilibrium periods* and (2) *revolutionary periods*. Equilibrium periods are relatively long and stable periods, in which the environment develops slowly and in a predictive manner. During these periods, teams are primary concentrated on advancing and optimizing existing procedures and products (i.e., incremental adaptation). It is important to mention though, contrary to the view of some authors (e.g., Brown & Eisenhardt, 1997), that the PEP does not exclude development and innovation from equilibrium periods (Tushman & Romanelli, 1985). It is rather assumed that these processes are incremental (cf. exploitation, March, 1991) during equilibrium periods.

Upheaval, however, in the form of sudden and disruptive events (punctuations), mark the beginning of revolutionary periods (Tushman & Romanelli, 1985). In these periods, teams

are required to fundamentally modify processes and strategies (cf. exploration, March, 1991). The driving forces behind revolutionary periods are internal or external shocks (cf. event system theory, Morgeson, Mitchell, & Liu, 2013). Internal shocks can be introduced for example by changes in top management (Romanelli & Tushman, 1994), or leadership interventions (Morgenson, 2005). In contrast to self-initialized interventions, external shocks threaten the way a system (e.g., a team) obtains its resources (Wollin 1999), and are more likely to disrupt an equilibrium than internal interventions (Humprey & Aime, 2014). They may result from fundamental changes in technology (Pulakos et al., 2000; Tushman & Rosenkopf, 1992; Zellmer-Bruhn, 2003), but also economic crisis (Romanelli & Tushman, 1994), or natural disasters (Tilcsik & Marquis, 2013).

4.3 Empirical Evidence on Punctuated Equilibrium

In the empirical work of punctuated equilibrium theory (PET), two different streams become apparent: a team and project specific ‘midpoint-transition version’ (Arrow, 1997) originating from Gersick’s (1989) research, and a more general version that focuses on team (and organizational) evolution over the long term in dynamic environments.

In the midpoint version, it is assumed that project teams develop relatively quickly towards a stable state (equilibrium). Thus, they modify their existing structures fundamentally after the midpoint of a project (Gersick, 1989), by initializing a brief revolutionary period on their own. This happens because teams perceive a discrepancy between the current and the desired status. This version of the PET received mixed empirical support. Whereas six of the eight (87.5%) teams in Gersick’s original study (1989) experienced midpoint transitions, Chang, Bordia, and Duck (2003) found that only nine out of 25 teams (36%) showed signs of such a transition. Arrow (1997) even demonstrated, that of the 20 examined teams in her study, none showed any signs of behaviors corresponding to the assumptions emerging from the midpoint-version PEP. Similarly, Lim and Murnighan (1994) reported in their analysis of

136 undergraduate students that they "found no consistent point where transitions occurred and no clearly identifiable transition period" (p. 166) (see for similar results Waller, Zellmer-Bruhn, & Giambastista, 2002). However, when re-analyzing video data from previous studies of 80 teams, Okhuysen and Waller (2002) reported that more teams showed signs of transitions, when they were explicitly instructed to use time management as part of their work process (compared to only 30% of the groups in the control condition). Research from Seers and Woodruff (1997) pointed in a similar direction, indicating that transitioning from equilibrium to revolution does rather reflect task pacing under a deadline. The deadline may become salient at the midpoint of a project, but might also be relevant during earlier or later team phases.

Research on punctuated equilibrium that moves beyond the midpoint question, is, however, rather scarce. A rare exception pertains to the work from Romanelli and Tushman (1994). In their examination of 25 computer producers, they found that most organizational transformations were indeed discontinuous and rapid, and that small changes did not accumulate to fundamental transformations.

It is apparent from the short overview of empirical studies, that the punctuated equilibrium domain in general does not provide extensive insights into the conditions or processes that facilitate or impede adaptation. Some inferences can be drawn at least from studies that did not explicitly examined elements of the PEP, but focused on individuals or teams adapting to non-routine or unexpected events that match the characteristics of punctuations. An overview of this research can be found in Table 1. However, despite many important findings, the majority of the research did not focus on the processes in adaptation (cf. Baard, Rench, & Kowslowski, 2014; Maynard, Kennedy, & Sommer, 2015), and was often based on relatively small samples.

Table 4.2: Overview of quantitative studies focusing on adaptation after punctuations

Authors	Type	Level	Sample	Task-type	Results
Burtscher et al. (2010)	Longitudinal	Team	22 anesthesia teams	Field	Higher performing teams showed more adaptive coordination after an unexpected event.
Burtscher et al. (2011)	Longitudinal	Team	15 anesthesia teams	Laboratory (Simulation)	Higher performing teams showed more adaptive coordination through information management after an unexpected event.
Lang & Bliese (2009)	Longitudinal	Individual	184 students	Laboratory (Military simulation)	GMA predicts the drop of performance after an unexpected event.
LePine (2003)	Longitudinal	Team	73 student teams	Laboratory (Military simulation)	GMA, predicts performance after an unexpected event (as well as dependability, achievement and openness). This was mediated by role structure adaptation.
LePine (2005)	Longitudinal	Team	64 student teams	Laboratory (Military simulation)	GMA predicts performance after an unexpected event. Teams with difficult goals and high learning goal orientations were likely to adapt, whereas teams with high performance goal orientations were unlikely to adapt.
LePine et al. (2000)	Longitudinal	Individual	75 students	Laboratory (Military simulation)	GMA, conscientiousness, and openness predicted performance after an unexpected event.
Morgeson (2005)	Cross-sectional	Team	29 teams from various domains	Field	Leader preparation and supportive coaching were more strongly related to effectiveness when the novelty of non-routine events was high. Moreover, active leader intervention activities were positively related to effectiveness after disruptive events.

Randall et al. (2011)	Longitudinal	Team	74 student teams	Laboratory (City-management simulation)	Team GMA predicted similarity and accuracy of team mental models, as well as information sharing, which predicted how well teams could adapt to a sudden change in managing a simulated city.
Stachowski et al. (2009)	Longitudinal	Team	14 nuclear power plant crews	Field (Power plant simulation)	Higher performing crews exhibited fewer, shorter, and less complex interaction patterns.
Waller (1999)	Longitudinal	Team	10 airline flight-crews	Field (B-727 flight simulator)	Longer reaction-times to prioritize tasks or distribute activities after nonroutine events negatively affects performance.
Waller et al. (2004)	Longitudinal	Team	14 nuclear power plant control room crews	Field (Power plant simulation)	Information collection, shared mental model development activities, and intra-crew processes used during model development, differ between lower- and higher-performing control crews during nonroutine tasks.
Zellmer-Bruhn (2003)	Cross-sectional	Team	90 sales and service teams	Field	Interruptions influence knowledge transfer effort and are positively related to new work routines.

To conclude, in spite of a considerable number of studies that examined the validity of the PEP and leveraged upon its principles, three major shortcomings can be identified in previous research. First, the theory provides a structural frame for change on a global level and ignores the processes that drive development and adaptation during equilibrium and revolutionary periods (Romanelli & Tushman, 1994; Sastry, 1997). This general lack of understanding for the underlying mechanism of adaptation (Baard et al., 2014; Maynard et al., 2015) is also reflected in the lack of punctuated equilibrium-based studies that examine the influence of team processes. Therefore, to find out behavioral patterns that are effective in revolutionary or equilibrium periods, we propose to integrate recent taxonomies of team processes into the PEP.

Secondly, psychological research often tends to lean towards a general 'the more the better' perspective. Thus, the effects of processes on the success of adaptation may differ, depending on the demands that result from revolutionary and equilibrium periods. A process that might support adaptation in revolutionary periods, could possible inhibit successful adaptation during equilibrium periods, and vice versa. Hence, we propose to integrate a contingency perspective into the theory, to capture how conditions and contexts affect process-outcome relations in punctuated equilibrium.

Thirdly, adaptation to change cannot be conceived as an isolated event, but as a reoccurring form of self-regulation (Carver & Scheier, 1998; DeShon, Kozlowski, Aaron, Schmidt, Milner, & Wiechmann, 2004). However, even though the punctuated equilibrium theory is understood as a theory of change and development, it focuses only on a macro-level of change, which does integrate the performance-feedback cycles that represent the notion of team adaptation (Humphrey & Aime, 2014). Adaptation in punctuated equilibrium should therefore be understood as process within an episodic system, in which relations to previous and subsequent states should be considered systematically. In the following, we will therefore

elaborate an enhanced version of the punctuated equilibrium theory, providing possible solutions to the elaborated shortcomings.

4.4 Advancing Punctuated Equilibrium Theory

To start with, we follow recent taxonomies of team processes that conceive adaptation to occur in performance episodes (Marks et al., 2001). Performance episodes are "distinguishable periods of time over which performance accrues and feedback is available" (p. 359). Importantly, outputs of one episode become inputs for the consecutive episode (IMOI models, see Ilgen et al., 2005). The resulting process-outcome cycles are determined by team activities that are sometimes directed at goal accomplishment, and sometimes at the evaluation of past behaviors (Marks et al., 2001). Therefore, the processes contributing to adaptation during a performance episode in the PEP can be categorized as processes taking place in *action-phases* and *transition-phases*. Action phases are defined as "periods of time when teams are engaged in acts that contribute directly to goal accomplishment (i.e., taskwork)" (Marks et al., p. 360). Transition phases are defined as "periods of time when teams focus primarily on evaluation and/or planning activities to guide their accomplishment of a team goal or objective" (p. 360).

Hence, to adapt, teams must "execute different processes at different times, depending on task demands" (Mathieu, Maynard, Rapp, & Gilson, 2008, p. 414). For instance, Hollenbeck, Moon, Ellis, West, Ilgen, Sheppard, et al. (2002) examined 80 four-person teams engaged in a military computer simulation and found, that advantages of high cognitive abilities can be neutralized when the structure of a team does not fit to the environmental demands. Similarly, Aime, Humphrey, Derue, and Paul (2014) reported that power structures in teams are dynamic and that power actively shifts among team members in accordance to the demands resulting from the current task. This contingency perspective (Lawrence & Lorsch, 1967) illustrates that there is rarely a strategy that is best suited for all scenarios.

Instead, strategies should be adjusted to fit the demands. However, as mentioned above, demands of revolutionary and equilibrium periods may differ, requiring the execution of different activates. In the punctuated equilibrium theory, activities are generally concerned with the formation and maintenance of so-called *deep structures* (Gersick, 1991). Deep structures can be understood as the underlying order of a system (Wollin, 1999). A deep structured system is typically characterized by complex, interconnected and interacting parts, including the basic activity patterns that provide the functioning and maintain its existence. It is comprised of "beliefs, values, culture, technology, operating routines, control systems, organizational structures (formal and informal), resources, core competencies and distributions of power that exist within an organizational system" (Wollin, 1999, p. 360). During revolutionary periods, teams must (re)build and develop a deep structure, by focusing on exploration-associated processes. This includes activities as for example search, risk taking, or experimentation (March, 1991). Revolutionary periods are concluded by the establishing of the deep structure, resulting in the initiation of the following equilibrium period. During these periods "the system's basic organization and activity patterns stay the same" (Gersick, 1991, p. 16). Teams therefore must focus on exploitation strategies (March, 1991) that are associated with optimization, refinement, and maintenance of the deep structure (see for reviews on exploration and exploitation Gupta, Smith, & Shalley, 2006; Li, Vanhaverbeke, & Schoenmakers, 2008). Hence, dominating processes during revolutionary periods should rather be directed at the development of deep structures, whereas processes during equilibrium periods should be directed at maintaining deep structures.

In sum, we propose to enhance Gersick's punctuated equilibrium theory with the episodic framework from Marks et al. (2001). We therefore integrate action-phase and transition-phase processes, which are nested in performance episode and develop in accordance with the underlying deep structure of a team (see Figure 4.1).

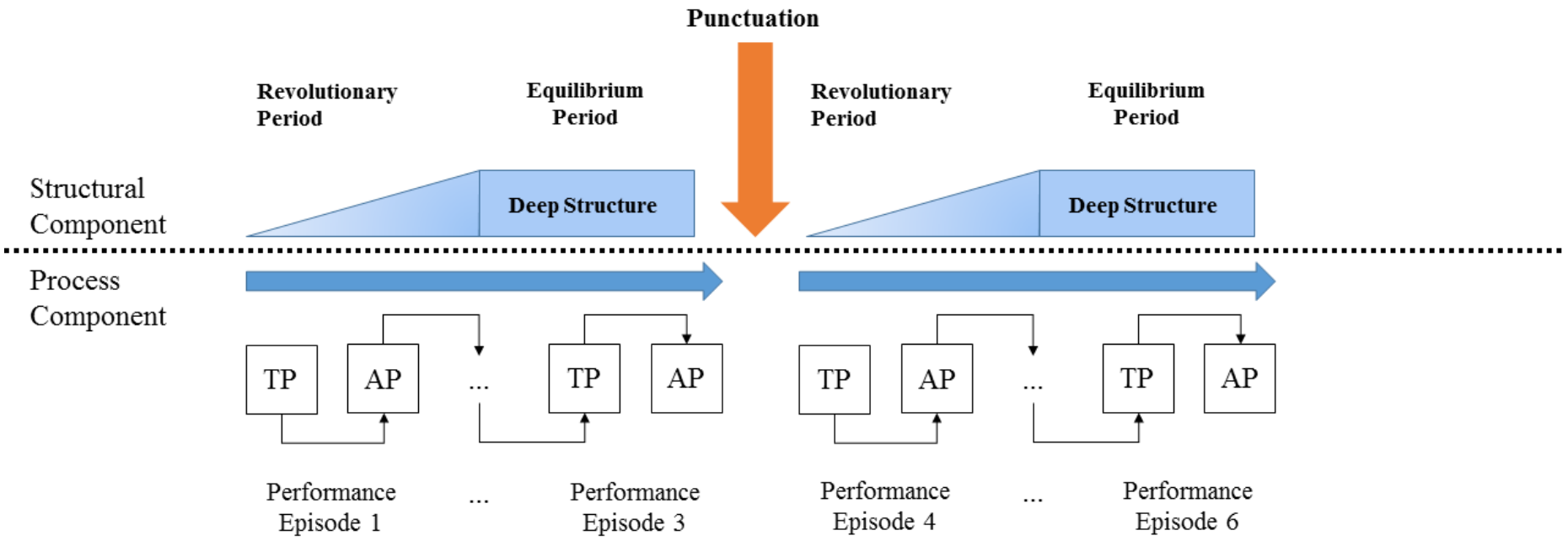

Figure 4.1: Advanced punctuated equilibrium model

Moreover, we propose a contingency perspective on punctuated equilibrium, and assume, that the relation between processes and outcomes is influenced by contextual factors that emerge from demands resulting from equilibrium and revolutionary periods.

4.5 Team Reflection in Punctuated Equilibrium

As mentioned before, research on adaptation with a punctuated equilibrium background has scarcely examined the influences of processes. Moreover, adaptation research in general (see for some examples Table 4.1) concentrated mainly on team characteristics (e.g., general mental ability, Lang and Bliese, 2009), and sometimes action-phase processes, as for example coordination and communication (Maynard et al., 2015). Relevant transition-phase processes are rarely examined.

A central team transition-phase process is *team reflection*. Zajac, Gregory, Bedwell, Kramer, and Salas (2014) for example argued that "adaptation is, essentially, a manifestation of how the team generalizes past knowledge and successfully transfers it to a new situation" (p. 63). LePine, Piccolo, Jackson, Mathieu, and Saul (2008) summarized, that transition-phase processes are generally concerned with reflection upon and interpreting previous team accomplishments. Marks et al. (2001) also argued that reflection (among planning) is a fundamental transition-phase process. In this study, we will therefore focus on the role of the transition-phase process of *team reflection* in punctuated equilibrium.

Team reflection can be defined as looking back on experiences, goals, actions, working methods, strategies, and assumptions (Savelsbergh, van der Heijden, & Poell, 2009). Team reflection may help teams to adapt, because decisions and new plans or strategies can be built upon experiences and knowledge. Moreover, team reflection allows the identification of successful and unsuccessful procedures and strategies and the elements that contribute to their effectiveness. In contrast, non-reflexive teams are determined by acting without the awareness of the action, a tendency to blindly react to situations, or by an unwillingness to

acknowledge and deal with errors (West, 2000).

Reflexive teams are in general assumed to show higher levels of performance (e.g., Konradt & Eckardt, 2016) and to be more innovative (e.g., Schippers, West, & Dawson, 2015). However, previous research did sometimes also reveal detrimental and neutral effects (Konradt, Otte, Schippers, & Steenfatt, 2016; Moreland & McMinn, 2010). Otte et al. (2017) argued that one reason for these diverse results is that team reflection measures captured only quantitative elements of reflection (i.e., frequencies of reflexive behaviors, as in 'how often'). This perspective indicates a relatively simple 'the more the better' relation between reflection and outcomes (e.g., performance), which could have led to mixed results. Teams might for example try to reflect upon too many different topics, which might hinder them to reach significant depth. Depth however, can be essential for finding the root of a problem. Gurtner Tschan, Semmer, and Nägele (2007) for example found that the 49 studied teams often did not reach substantial depth in their discussions. Instead, they focused on general, rather than task-specific strategies, which added noise, instead of helpful communication. Thus, it is worth mentioning that qualitative reflection should not be conceived as another 'the more the better' dimension. As Konradt et al. (2016) argued, one dimension might come at the expense of the other. Extremely detailed or thorough reflection focused at an isolated topic, might consume too much time and keep teams from elaborating other, relevant topics (see also Schippers, Homan, & van Knippenberg, 2013).

4.5.1 A Contingency Perspective on Reflection

Another reason for the aforementioned mixed results might be a more complex relation between reflection and performance. Since time is a limited resource, teams might generally not be able to cover all topics (quantity) with substantial depth (quality) (Otte et al., 2017). A possible key-factor in reflection research could therefore be the focus on the appropriate reflexive behaviors, depending on the demands imposed by the task or the

environment. Schippers et al. (2015) for example found in a field study of 98 primary health care teams, that the interaction of reflection and work demands predicted innovation. Teams working in a low-quality physical work environment showed high benefits of reflection, whereas this relation was low in high-quality environments. Similarly, Otte et al. (2017, Study 4) found in a sample of 62 students learning for an exam, that reflection is most effective when the reflexive focus is adjusted (e.g., focus on the qualitative or quantitative reflexive processes) dynamically per demands. These important results indicate that, in order to understand the effects of reflection on performance, one must abandon a 'one size fits all' solution. Following the contingency perspective on the effectiveness of processes in punctuated equilibrium, we argue that a contingency approach to team reflection (Schippers et al., 2015) is indispensable to not only explain the complex team reflection-performance relationship, but to also advance team reflection theory and research, in order to draw more elaborate conclusions and eventually, provide better support for the development of practical oriented interventions and trainings.

Hence, we argue that teams must focus on the appropriate type of reflexive behaviors, following the situational demands resulting from revolutionary or equilibrium periods. A fit between reflexive behaviors and demands is beneficial for the performance of an organizational unit (cf. Tushman & Romanelli, 1985). However, units that exhibit a miss-fit between demands and reflexive focus, might yield neutral, or even detrimental effects of reflection on performance (cf. Hollenbeck et al., 2002), because they cannot distribute their resources effectively, come to wrong conclusions, and may be generally affected by biased information processing.

4.5.2 Team Reflection and Adaptation in Revolutionary Periods

Revolutionary periods are characterized by fundamental changes (punctuations) that can cause disruptions within a team (or an organization). Most importantly, punctuations

require teams to adapt by rebuilding their deep structure. We therefore argue that team reflection can help teams to identify the relevant elements of the deep structure, and consequently, plan and implement necessary changes. A team will only benefit from team reflection, when the reflexive process fits the environmental demands. In revolutionary periods, we therefore propose that teams have to engage in the explorative process of qualitative reflection (deep and thorough reflection). This may help to identify and elaborate the key elements of the deep structure and to extract relevant information that allows to introduce appropriate changes.

Hypothesis 1: After a punctuation, qualitative team reflection is positively related to team performance.

Consequently, teams that focus on superficial reflection (i.e., quantitative reflection) are not able to determine key elements and may leave them therefore either unchanged, or even build their future strategies upon biased conclusions, resulting in changes that are detrimental for the team.

Hypothesis 2: After a punctuation, quantitative team reflection is negatively related to team performance.

4.5.3 Team Reflection and Adaptation in Equilibrium Periods

Equilibrium periods are characterized by incremental progress, because no fundamental changes are made to the deep structure of the system during this period. During equilibrium periods, teams mainly show routine behaviors (Gersick, 1989; Zellmer-Bruhn, 2003) and optimize procedures in the form of minor adjustments (Gersick, 1991) to the deep structure. Hence, we argue that optimization activities do not require to question or change fundamental principles. However, as products become more complex and elaborate over time, available resources have to be distributed carefully. Tushman and Romanelli (1985) for example proposed that the greater the structural complexity and interdependence, the greater

the emphasis in incremental development. Therefore, quantitatively focused team reflection is conceived as exploitative team process, which becomes particularly important during equilibrium episodes, because it allows covering all relevant topics.

Hypothesis 3: During equilibrium periods, quantitative team reflection is positively associated with performance.

Since ongoing optimization of the deep structure will increase complexity of the whole system, identifying (and consequently changing) elements of a deep structure via qualitative reflection might over time become more difficult, challenging, and consume more resources. This might therefore render successful changes highly unlikely. Moreover, attempts to adjust the deep structure may lead to overcompensation, because the environmental reactions to team outputs (e.g., strategy adjustments) are not instantaneous (cf. lag time, see Carver & Scheier, 1998). As a result, mal-adaptation might occur. Also, teams that still deeply reflect during equilibrium periods might experience difficulties finding successful strategies on a more basic level (cf. Lang & Bliese, 2009). Hence, even though these teams did establish a deep structure, it does not help to successfully compete for resources (Wollin, 1999). Therefore, we argue that qualitative team reflection during equilibrium periods has detrimental effects on performance.

Hypothesis 4: During equilibrium periods, qualitative team reflection is negatively associated with performance.

4.6 Method

4.6.1 Participants and Setting

We recruited overall 352 graduate students (41.8 % female) studying business management at a German university of applied science. The participants were randomly assigned to 86 teams (ranging between three and five members; $M = 4.09$, $SD = 0.93$). Participant's mean age was 24.89 years ($SD = 2.76$) and they had an average working

experience of 2.7 years (*SD* = 2.48). Data collection started in winter of 2014 and ended in winter of 2016. Since we started the data collection one semester earlier for some measurement points, and 3 teams dropped out of the business game during the final business years, we ended up with 69 teams for the first three measurement points, 86 for the fourth, 84 for the fifth, and 83 for the sixth measurement point.

The teams participated in the business simulation TopSim General Management II (TATA Interactive Systems GmbH, Mumbai, India), as part of their management course that has been used in other studies (Konradt & Eckardt, 2016). Their task was to manage a fictional enterprise in the production sector over the course of multiple business years. To maximize the company's profit, teams had to make complex and conflicting decisions in relevant business operations. Moreover, the companies operated in a dynamic marked environment of a maximum of ten companies. In each marked, the decisions of a single company affected the decisions of all competitors within the respective market. Teams therefore had to constantly develop and adapt their strategies to remain competitive. The marked environments were furthermore shaped by external influences that were simulated by the game, as for example economic growth, or the accessibility of new markets.

One business year took teams one week to complete, after which they received feedback on their performance, and started working on their strategies for the next business year (cf. IMOI models, Ilgen et al., 2005). The feedback also included information on the expected economic conditions teams had to face during the next business year.

4.6.2 Measures

Team Reflection. Team Reflection was measured with the REMINT (Reflection measure for individuals and teams, Otte et al., 2017). The scale represents two reflexive behaviors. That is *information seeking* ("gathering information from various sources for returning unprocessed information on the status of the team", p. 3), and *information*

evaluation (“aggregating pieces of information and making interpretations that go beyond their individual significance”, p. 3). Both dimensions exist as qualitative and a quantitative dimension, resulting in four dimensions, with four items each. A sample item of quantitative information seeking was ‘we collected information about our successes and failures very frequently’. A sample item of qualitative information evaluation was ‘we made very detailed evaluations of the quality of our work’. The response format was a five-point Likert scale ranging from *strongly disagree* (1) to *strongly agree* (5). All four dimensions were repeatedly measured after each business year. Reliabilities are depicted in Tables 4.2 and 4.3. The calculated $r_{WG(J)}$ values (Lindell, Brandt, & Whitney, 1999) ranged between 0.63 and 0.76, which can be judged as acceptable to good within-team agreement (Cohen, Doveh, & Nahum-Shani, 2009). Responses were therefore aggregated to the team level (see also Tables 4.2 and 4.3).

Team Performance. Team performance was measured as each company’s net profit after each business year. The value resulted from several parameters (e.g., units sold, price of the product, marketing budget, or number of employees) that teams had to plan and adjust during each business year and was generated by the simulation software. The net profit is a good performance indicator for two reasons. First, from an economic point of view, it is the main basis for any kind of indicator for efficiency and productivity of a business, e.g., the return on investment, the return on sales, etc. Furthermore, the net profit is the prerequisite for keeping and improving the equity, and in this way, the liquidity of the company. Second, based on this argumentation also for the participants the net profit regularly used as the main indicator of success. The experiences from a huge number of seminars shows that the first look of participants who get the business report always goes to the profit and loss statement, and there the bottom line which shows the net profit. The mean net profit values for each business year and standard deviation are depicted in Tables 4.2 and 4.3.

4.6.3 Design and Procedures

We used a six-wave panel design, measuring performance, information seeking (quality and quantity), and information evaluation (quality and quantity), after each business year. Prior to the data collections, participants were informed about the study's context and were invited to participate. In addition to measuring reflection, we assessed demographic variables during the first business year. Furthermore, each participant received an individual identification code that allowed us to match the data across all six waves. Participants were informed that this affected anonymity.

4.6.4 Punctuation, equilibrium and assignment of hypothesis

We conceived one business year as one performance episode (Marks et al., 2001). The business simulation started with a quasi-revolutionary period, because teams had to develop a new deep structure by developing and implementing a strategy they wanted to pursue. We therefore assumed that the first measurement point can be conceived as revolutionary period (see hypotheses one and two). We then conceived the second and third measurement points as equilibrium period, because teams should have established a deep structure by then (i.e., chosen a strategy and created a structure for their fictional organization) and should rather be focused on optimizing procedures (see hypotheses three and four).

After the third measurement point, teams had to adapt to a punctuation in the form of an economic crisis. As described in Romanelli and Tushman (1994), economic crisis can result in punctuations and introduce revolutionary periods. We therefore expected the crisis to be strong enough to require teams to adapt their deep structures and focus again on exploration rather than exploitation. Hence, over the course of the fourth, fifth and sixth measurement point, we expected a replication of the pattern from measurement points one to three. The relations of quality and quantity of reflection and performance, as predicted by our hypothesis, are depicted in Figure 4.2.

4.6.5 Data Analysis Strategy

We used cross-lagged panel analysis based upon structural equation modeling for examining the longitudinal structure in our dataset. However, accounting for the complexity of our models, the sample size on the team level (N = 86) may produce biased estimates when using maximum likelihood estimations (cf. Hinkin, 1998). We therefore used Bayesian structural equation modeling (BSEM) (see Muthén & Asparouhov, 2012; Zyphur & Oswald, 2015). BSEM has several advantages over common maximum likelihood estimations, including that it allows for less-biased estimates in smaller sample sizes (Song & Lee, 2012). In order to provide stable estimations of our models, we employed two Markov chain Monte Carlo chains and set the number of iterations to 100,000 (cf. Zypher & Oswald, 2015). To assess model fit, we used the posterior predictive p-value (PPP) and the posterior predictive checking (PPC) 95% credibility interval (CI), which indicates whether the data fits the proposed model. A PPP value above .05 indicates an acceptable fit, whereas a PPP value of .5 indicates an excellent fit (Muthén & Asparouhov, 2012). The PPC CI should include a negative lower limit and a positive upper limit (Zyphur & Oswald, 2015) to indicate a good fit. We used the BIC (Bayesian information criterion) and the DIC (Deviance information criterion) for model comparisons (cf. Kaplan & Depaoli, 2012). A benefit of using the BIC and DIC for model comparison is that models do not need to be nested (Kaplan & Depaoli, 2012). In general, a lower BIC and DIC can be interpreted as model that has more likely to have generated the data (West, Taylor, & Wu, 2012). Finally, we verified whether the solution was stable when we increased the number of iterations. The analysis was conducted in Mplus 7.4 (Muthén & Muthén, 2017).

Moreover, to reduce overall model complexity, we decided to examine the effects of information seeking and information evaluation on performance separately. Also, as we expected the punctuation between T3 and T4 to impact the teams' deep structures, and to

further reduce model complexity, we separated our data into a pre-punctuation model (T1 to T3) and a post-punctuation model (T4-T6). Altogether, this resulted in four models for testing our hypotheses.

In the next step, as performance at a given point in time was dependent on the preceding performance (except for the first and the fourth measurement point), we set autoregressive relationships for the performance variables. This can also be interpreted as controlling for previous performance, or capturing the change in performance (i.e., performance *improvement*). However, we did not include the autoregressive relationships for the team reflection parameters, because this would have meant that we conceived the change in performance as dependent on the *change* in team reflection. Furthermore, based on recent findings in Konradt and Eckardt (2016), who reported significant lag-2 effects of reflection on subsequent performance (i.e., effects of preceding reflection that were significant above the direct effects reflection at given time point), we extended the model and allowed the same relations in our models. The final version of the four proposed models including all four hypotheses is depicted in Figure 4.2. We compared the model depicted in Figure 4.2 to two other variants. Since we excluded an effect of team reflection at a given time point on subsequent reflection, these relationships where implicitly set to zero.

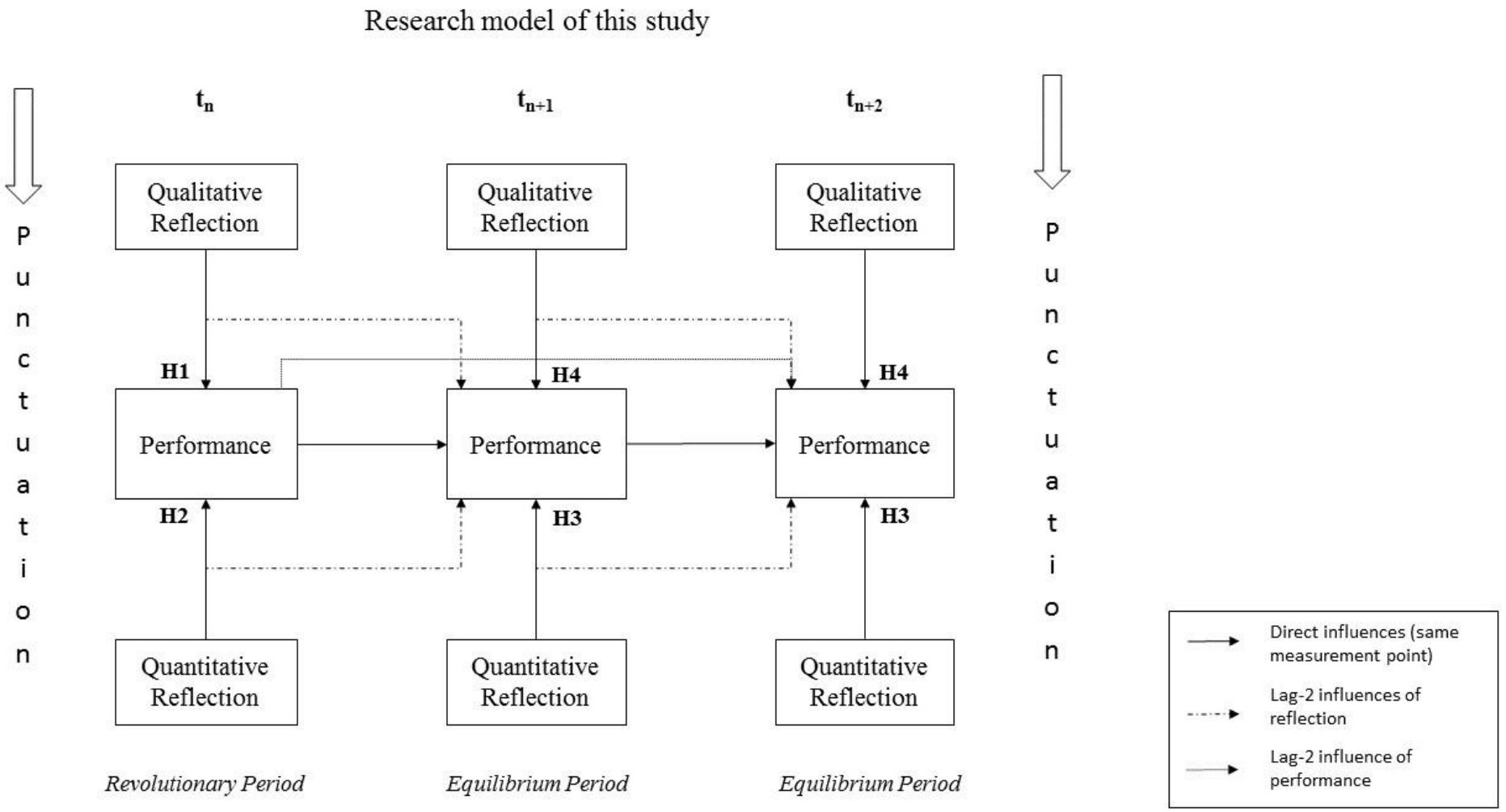

Figure 4.2: Punctuated equilibrium model and hypotheses

We therefore compared our proposed model (variant A) with a model with covariances (variant B), and a model with autoregressive effects (variant C) between team reflection measured at consecutive time points, in order to rule out that these alternative models fitted the data better.

Lastly, since our sample size was fairly small, compared to the complexity of our models, we set the significance level to .1, in order to be more sensitive regarding possible effects and to reduce the type II error rate.

4.7 Results

Descriptive statistics, reliabilities alpha, and the correlations between study variables can be found in Table 4.3 (T1 to T3) and Table 4.3 (T4 to T6). Model fit indices for all models, including variants A, B, and C, are depicted in Table 4.4.

The manipulation of the business simulation in the form of an economic crisis appears to have successfully initialized a revolutionary period, as the mean performance of all teams decreased significantly ($t(83) = 5.44$, $p < .001$) from T3 (5.94) to T4 (-0.22). Moreover, the punctuation appears to have been that extensive, that mean team performance continued to decrease over the following measurement points (T5 and T6), which indicates that the teams struggled to adapt to the change. Regarding the model fit, the initially proposed variant A, including only autoregressive relations between the performance variables, and no covariances or auto-regressive effects between the exogenous variables, fitted the data generally better than variants B and C, as indicated by the BIC and DIC values (see Table 4.4). Moreover, the PPP values (ranging from .304 to .569) and PPP confidence intervals for the four models (i.e., information seeking and information evaluation T1-T3 and T4-T6) based on variant A, can be judged as good to excellent.

Table 4.3: Correlation table of study variables of measurement points T1 – T3

	M (SD)	RWG(J)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Time 1																	
1 IS Qn.	4.02 (.50)	.70	**.90**														
2 IS Ql.	3.89 (.62)	.68	.83**	**.95**													
3 IE Qn.	3.75 (.54)	.65	.80**	.82**	**.93**												
4 IE Ql.	3.67 (.60)	.66	.75**	.85**	.83**	**.96**											
5 Perf.	3.16 (5.29)	--	-.02	.16	.09	.17	**NA**										
Time 2																	
6 IS Qn.	4.13 (.40)	.76	.71**	.66**	.73**	.65**	0.19	**.86**									
7 IS Ql.	4.04 (.44)	.73	.68**	.71**	.69**	.70**	0.08	.80**	**.88**								
8 IE Qn.	3.93 (.42)	.68	.68**	.66**	.73**	.61**	0.15	.82**	.76**	**.86**							
9 IE Ql.	3.89 (.47)	.68	.70**	.78**	.74**	.75**	0.09	.78**	.85**	.85**	**.92**						
10 Perf.	-.28 (9.56)	--	.01	.12	.02	.01	.20	.12	.13	.15	.19	**NA**					
Time 3																	
11 IS Qn.	4.02 (.39)	.66	.55**	.51**	.53**	.47**	0.03	.64**	.67**	.60**	.60**	-0.01	**.87**				
12 IS Ql.	4.01 (.42)	.70	.62**	.71**	.69**	.70**	0.10	.66**	.79**	.68**	.79**	0.12	.69**	**.89**			
13 IE Qn.	3.96 (.38)	.66	.54**	.54**	.59**	.53**	0.22	.62**	.63**	.64**	.66**	.31**	.72**	.71**	**.84**		
14 IE Ql.	3.96 (.43)	.68	.58**	.57**	.67**	.69**	0.22	.72**	.70**	.63**	.73**	0.15	.64**	.79**	.79**	**.92**	
15 Perf.	5.94 (12.76)	--	.01	.07	.06	.14	.38**	.06	.07	.09	.09	.18	.13	.25*	.13	.12	**NA**

Note. $N_{T1\text{-}T3}$ = 69. IS = Information seeking. IE = Information evaluation. Perf = Performance. Qn = Quantity. Ql = Quality. Reliability alphas on the diagonal (bold). All values are on team level.

*p < .05. **p < .01 (two-tailed).

Table 4.3: Correlation table of study variables of measurement points T4 – T6

	M (SD)	RWG(J)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Time 4																	
1 IS Qn.	4.00 (.47)	.68	**.85**														
2 IS Ql.	3.91 (.49)	.68	.76**	**.91**													
3 IE Qn.	3.92 (.45)	.66	.83**	.72**	**.89**												
4 IE Ql.	3.84 (.50)	.66	.67**	.82**	.77**	**.92**											
5 Perf.	-.22 (13.40)	--	.13	.19	.13	.20	--										
Time 5																	
6 IS Qn.	4.01 (.44)	.65	.84**	.72**	.78**	.63**	.17	**.90**									
7 IS Ql.	3.95 (.51)	.64	.79**	.80**	.77**	.73**	.28*	.83**	**.94**								
8 IE Qn.	3.91 (.46)	.63	.71**	.62**	.80**	.64**	.13	.84**	.78**	**.91**							
9 IE Ql.	3.90 (.43)	.63	.64**	.68**	.75**	.74**	.25*	.72**	.84**	.84**	**.92**						
10 Perf.	-4.36 (13.28)	--	.26*	.19	.23*	.29**	.44**	.31**	.34**	-31**	.25*	**--**					
Time 6																	
11 IS Qn.	4.04 (.43)	.69	.66**	.51**	.73**	.59**	.17	.70**	.53**	.67**	.57**	.160	**.90**				
12 IS Ql.	4.02 (.45)	.68	.68**	.76**	.80**	.76**	.20	.64**	.75**	.70**	.76**	.232*	.70**	**.91**			
13 IE Qn.	3.95 (.44)	.67	.70**	.50**	.75**	.61**	.14	.67**	.62**	.71**	.66**	.20	.85**	.77**	**.89**		
14 IE Ql.	3.94 (.47)	.68	.62**	.60**	.72**	.73**	.19	.56**	.64**	.65**	.70**	.15	.71**	.82**	.86**	**.91**	
15 Perf.	-7.97 (22.06)	--	.18	.15	.14	.20	.21	.10	.16	.12	.06	.44**	.17	.17	.14	.16	--

Note. N_{T4} = 86. N_{T5} = 84. N_{T6} = 83. IS = Information seeking. IE = Information evaluation. Perf = Performance. Qn = Quantity. Ql = Quality.

Reliability alphas on the diagonal (bold). All values are on team level.

*p < .05. **p < .01 (two-tailed).

The solution remained stable when further increasing the number of iterations above 100,000. Based on these results, and the rule provided by Occam's razor, we decided test our hypothesis with the simpler model excluding relations between the exogenous variables (Figure 4.2).

Table 4.4: Model comparisons

Model 1 – IS (T1- T3)	PPP	PPP CI	BIC	DIC
Model 1a: No covariances	.480	-20.201; 20.344	1505.793	1462.762
Model 1b: All covariances	.379	-25.64; 35.737	2197.612	2075.385
Model 1c: Auto-regressive paths	.080	-8.268; 52.818	2188.777	2093.488
Model 2 – IE (T1-T3)				
Model 2a: No covariances	.304	-15.203; 25.203	1514.280	1471.255
Model 2b: All covariances	.306	-22.604; 39.395	2189.875	2067.527
Model 2c: Auto-regressive paths	.099	-10.881; 52.177	2176.790	2081.458
Model 3 – IS (T4-T6)				
Model 3a: No covariances	.335	-16.026: 24.038	2055.865	2009.729
Model 3b: All covariances	.315	-23.337; 38.341	2385.904	2265.555
Model 3c: Auto-regressive paths	.002	13.979; 74.873	2397.750	2303.133
Model 4 – IE (T4-T6)				
Model 4a: No covariances	.569	-22.146; 18.380	2053.069	2006.944
Model 4b: All covariances	.425	-27.573; 33.827	2363.661	2243.494
Model 4c: Auto-regressive paths	.001	18.742; 79.931	2385.641	2290.979

Note. IS = Information seeking. IE = Information evaluation. PPP = posterior predictive p-value. PPP CI = Posterior predictive checking confidence interval. BIC = Bayesian information criterion. DIC = Deviance information criterion. Model variant A: No relations between exogenous variables. Model variant B: Covariance-based relations between exogenous variables. Model variant C: Auto-regressive path between exogenous variables.

4.7.1 Hypothesis Testing

The results of the analysis of the four models are depicted in Figures 4.3 to 4.6. An overview of confirmed and rejected hypotheses can be found in Table 4.5. An overview of the path estimates, posterior S.D.s, one-tailed p-values, as well as 90% confidence intervals for models 1 and 2 can be found in Appendix B, and for models 3 and 4 in Appendix C.

Hypothesis 1, which suggested that qualitative reflection during revolutionary periods is positively related to performance (controlled for quantitative reflection), was mainly confirmed. Three out of four relationships were significantly related to performance (Model 1: $\beta = 0.545$, posterior S.D. = 0.190, CI [0.206 | 0.831], $p < 0.01$; Model 2: $\beta = 0.274$, posterior S.D. = 0.204, CI [-0.077 | 0.594], $p < .1$; Model 4: $\beta = 0.299$, posterior S.D. = 0.163, CI [0.017 | 0.554], $p < .05$). Only quality of information seeking in the third model was not significantly related to performance, but the estimate pointed into the right direction.

Hypothesis 2, which predicted a negative relation between quantitative reflection and performance during revolutionary periods (controlled for qualitative reflection), was for the most part rejected. Only quantitative information seeking in Model 1 ($\beta = -0.474$, posterior S.D. = 0.193, CI [-0.768 | -0.133], $p < .05$) had a significant negative relation to performance.

Hypothesis 3 predicted a positive relation between quantitative reflection and performance during equilibrium periods (controlled for qualitative reflection). This hypothesis was rejected for the second and third episode, but was partly confirmed for the fifth and sixth measurement point. Quantity of information evaluation (Model 4) was significantly positively related to performance at T5 ($\beta = 0.522$, posterior S.D. = 0.189, CI [0.193 | 0.814], $p < .01$), and quantity of information seeking (Model 3) was significantly positively related to performance at T6 ($\beta = 0.244$, posterior S.D. = 0.164, CI [-0.036 | 0.504], $p < .1$).
Finally, Hypothesis 4 assumed a negative relation between qualitative reflection and performance during equilibrium periods (controlled for quantitative reflection).

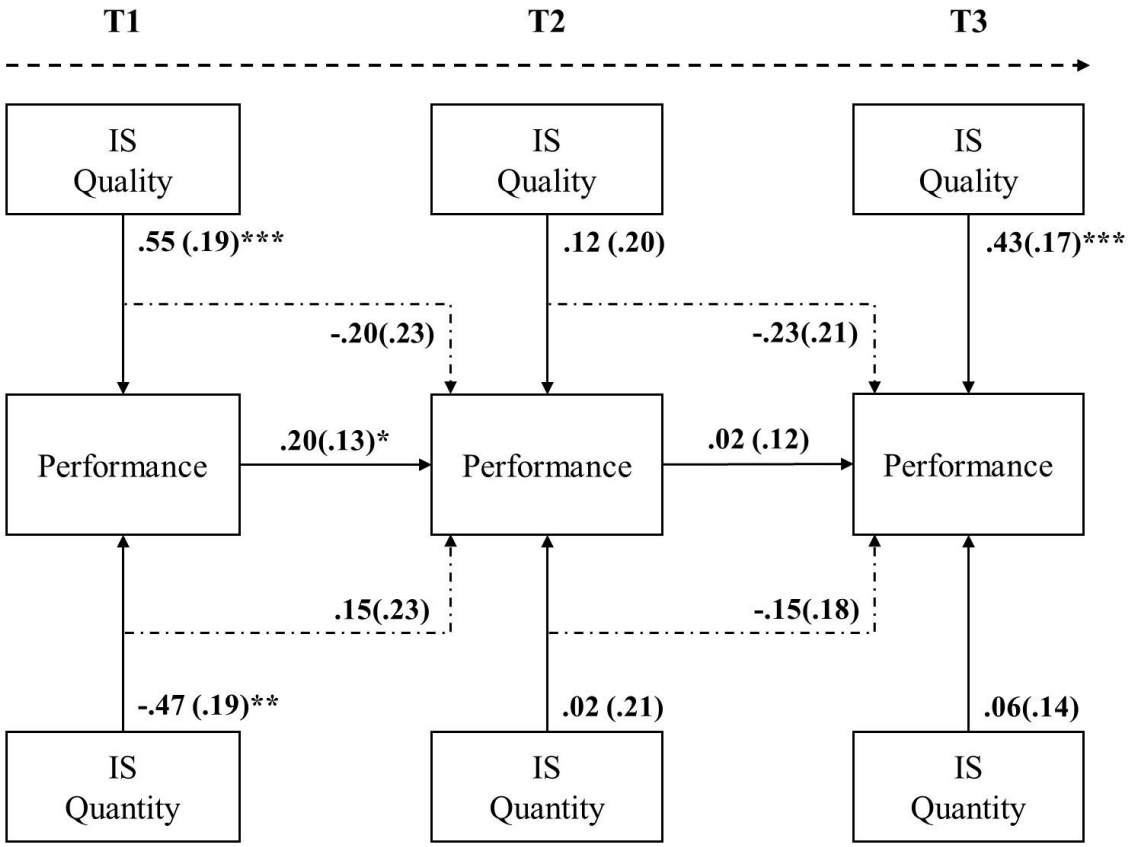

Figure 4.3: Model 1 including measurement points T1 to T3

Note. N_{Team} = 69. IS = Information seeking. All estimates are standardized. Standard errors in parenthesis. Lag-2 effect of performance T3 on T1 = 0.34 (0.11)***.

*p < .1. **p < .05. ***p < .01 (one-tailed).

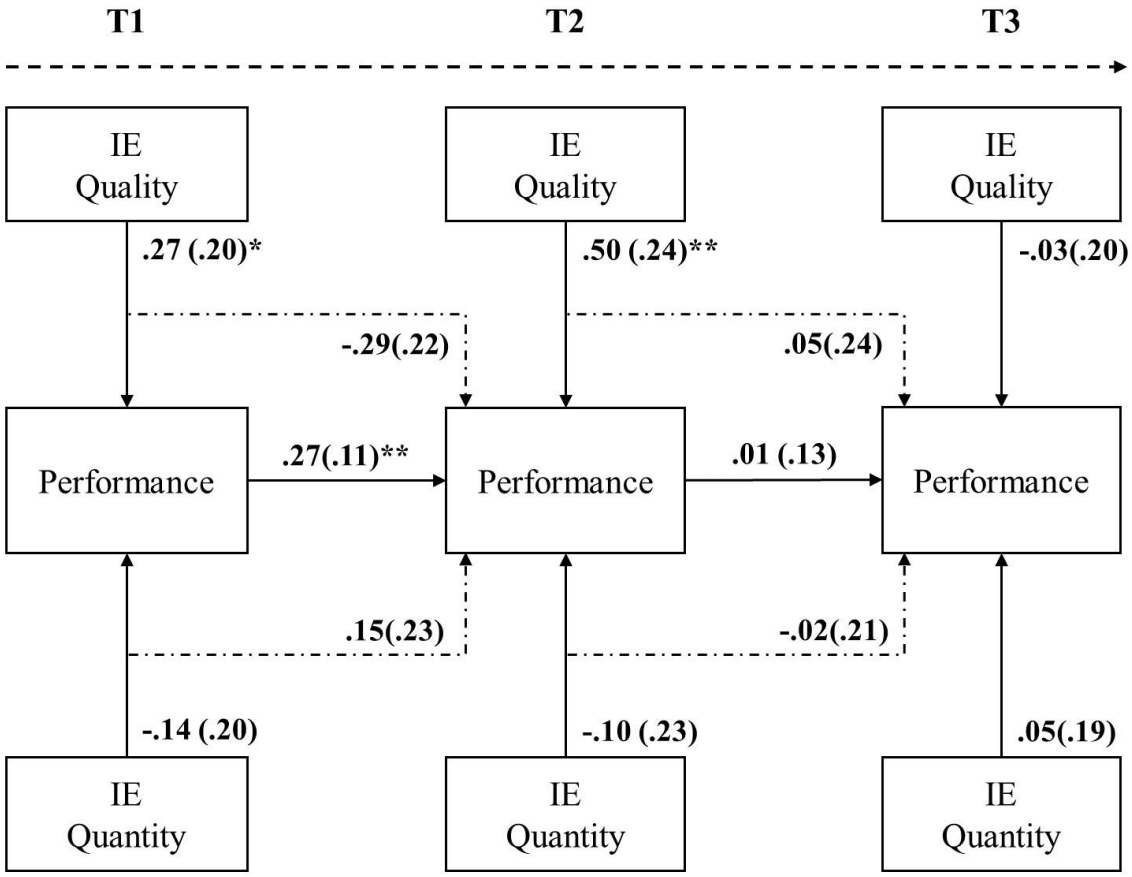

Figure 4.4: Model 2 including measurement points T1 to T3

Note. N_{Team} = 69. IS = Information seeking. All estimates are standardized. Standard errors in parenthesis. Lag-2 effect of performance T3 on T1 = 0.31 (0.12)***.

*p < .1. **p < .05. ***p < .01 (one-tailed).

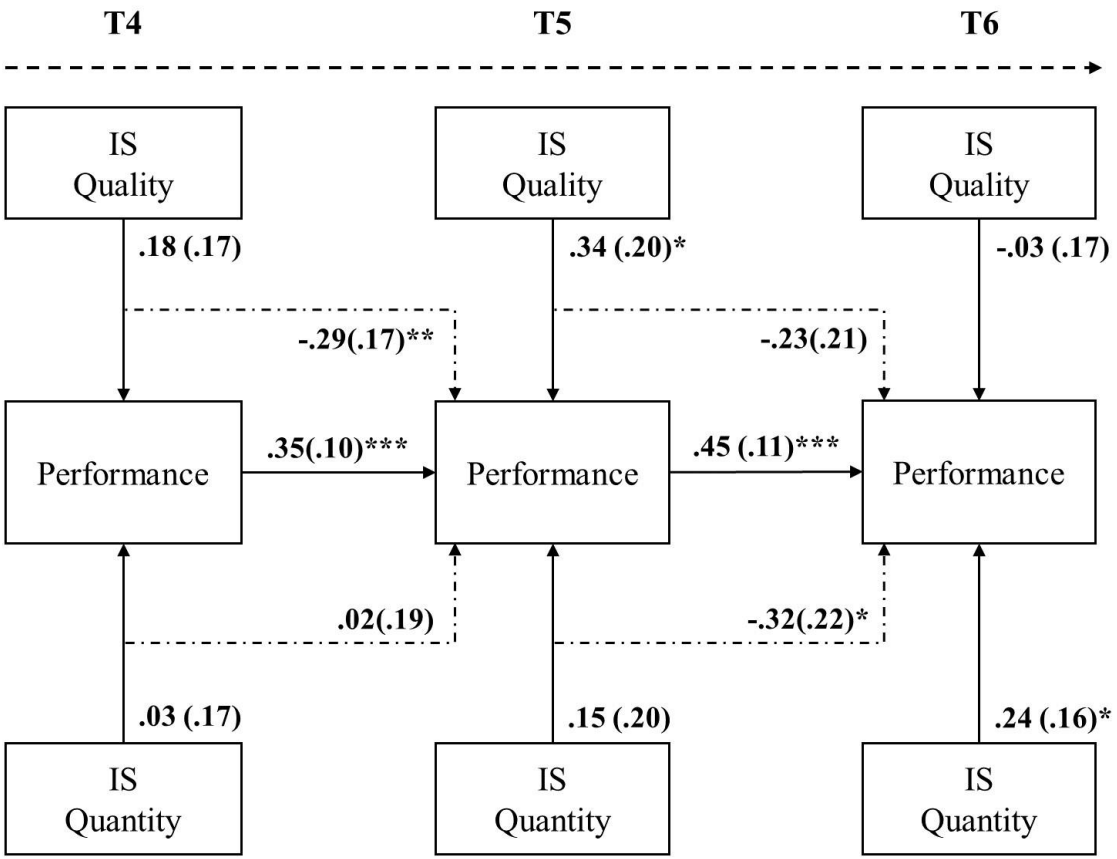

Figure 4.5: Model 3 including measurement points T4 to T6

Note. N_{Team} = 83-86. IS = Information seeking. All estimates are standardized. Standard errors in parenthesis. Lag-2 effect of performance T6 on T4 = -0.02 (0.12).

*p < .1. **p < .05. ***p < .01 (one-tailed).

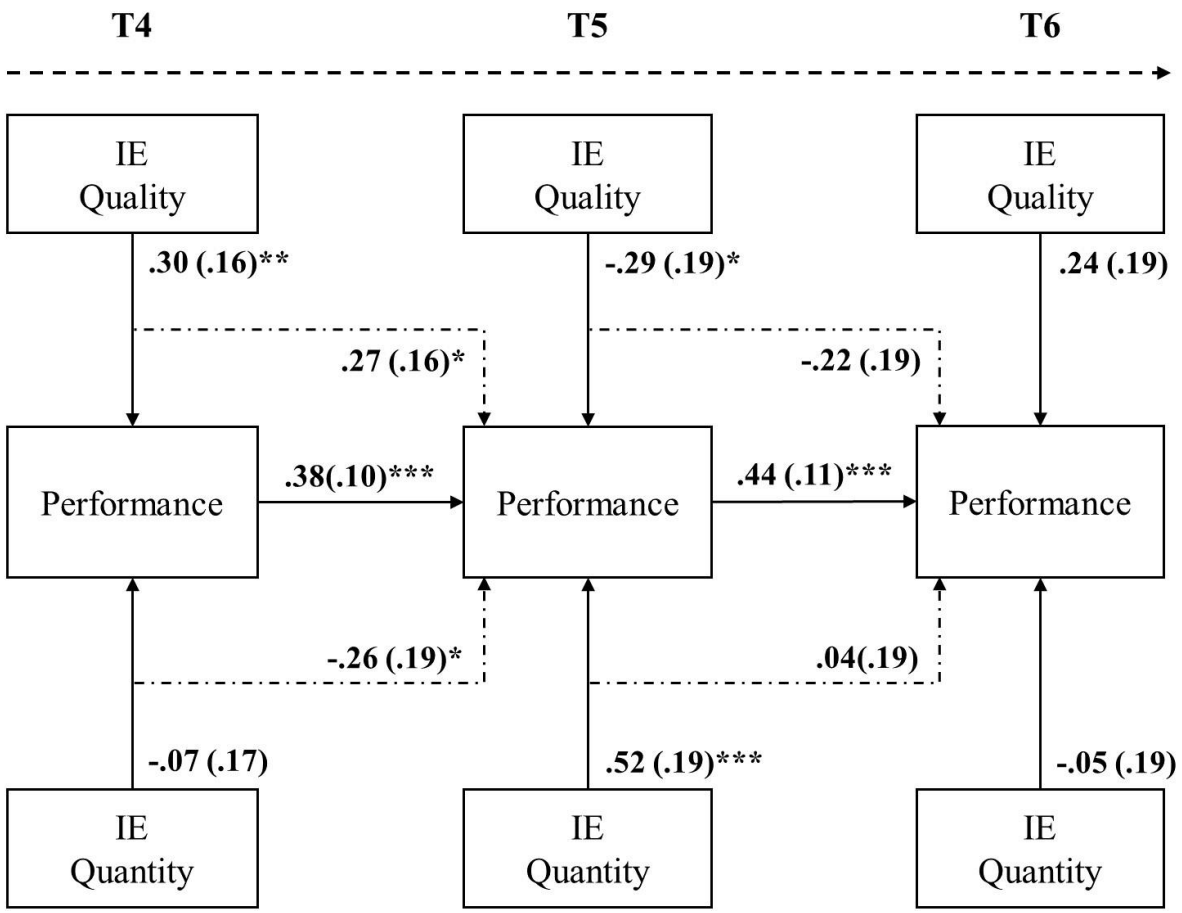

Figure 4.6: Model 4 including measurement points T4 to T6

Note. N_{Team} = 83-86. IS = Information seeking. All estimates are standardized. Standard errors in parenthesis. Lag-2 effect of performance T3 on T1 = 0.04 (0.12).

*p < .1. **p < .05. ***p < .01 (one-tailed).

This hypothesis was mainly rejected, as only qualitative information evaluation (Model 4) at T5 was significantly negatively related to performance (β = -0.286, posterior S.D. = 0.189, CI [-0.585| 0.035], $p < .1$). Moreover, qualitative information seeking in Model 1 at T3 (β = 0.433, posterior S.D. = 0.172, CI [0.136 | 0.700], $p < .01$) and Model 3 at T5 (β = 0.335, posterior S.D. = 0.202, CI [-0.002 | 0.664], $p < .1$), as well as qualitative information evaluation in Model 2 (β = 0.497, posterior S.D. = 0.235, CI [0.086 | 0.859], $p < .05$) were positively related to performance, even though these measurement points were assumed to be equilibrium periods.

Table 4.5: Overview of Hypotheses Testing

	H1	H2	H3	H4
Model 1 (IS T1-T3)	+	+	–	–
Model 2 (IE T1-T3)	+	0	–	–
Model 3 (IS T4-T6)	0	–	+	–
Model 4 (IE T4-T6)	+	–	+	+

Note. IS = Information seeking. IE = Information evaluation. H1-H4 = Hypothesis 1 to hypothesis 4. + = Hypothesis confirmed. – = Hypothesis rejected. 0 = Hypothesis partly confirmed.

4.7.2 Additional Analysis

Additionally, although we did not develop any hypothesis for them, several lag-2 reflection-performance relations significantly explained variance above the direct effects of reflection at the given time point. Most interestingly, we found in Model 4 (see Figure 4.6), that quality (β = 0.271, posterior S.D. = 0.161, CI [-0.002 | 0.528], $p < .1$) and quantity (β = -0.262, posterior S.D. = 0.188, CI [-0.564 | 0.055], $p < .1$) of information evaluation at T4 were significantly related to performance at T5 (controlled for quality and quantity of information evaluation at T5).

4.8 Discussion

In this research, we proposed and tested an advanced punctuated equilibrium theory. This theory proposes that different demands emerge from revolutionary and equilibrium periods, and that teams are required to adapt to these demands, by focusing on the relevant exploration or exploitation processes. Using longitudinal analysis, we tested several hypotheses that predicted how the explorative processes of qualitative team reflection, and the exploitative process of quantitative team reflection, were related to performance in this advanced punctuated equilibrium model.

First, we found a positive relation between qualitative reflection and performance during revolutionary periods. Interestingly, the relation between quality of information seeking/evaluation and performance was still significant at the second (information evaluation) and third (information seeking) measurement point (see Figure 4.3 and Figure 4.4). This finding may indicate that we have either incomplete assumptions regarding the relations between qualitative processes and performance, or that the revolutionary period did not end after the first measurement point, but continued at least throughout T2, or even T3. The assumption of an ongoing revolutionary period is reinforced by the fact that performance from T1 to T2 dropped instead of the expected increase, and that no quantitative team reflection relation to performance became significant during the first three measurement points. Additionally, we did not find the same pattern in T5 or T6 (i.e., Figures 4.5 and 4.6). A possible explanation would be that the complexity of the business simulation delayed the establishment of each team's deep structure.

The second hypothesis predicted a negative relation between quantitative reflection and performance during revolutionary periods. Contrary to the expectations, we did not find the predicted relations, as only quantitative information seeking at the first measurement point was negatively related to performance. However, neither could we replicate this finding after

the punctuation at T4, nor during the potentially fading revolutionary period at T2 and T3. Thus, regardless of the non-significant relations, the results did not, at any point, indicate a positive relation between quantitative reflection and performance during revolutionary periods. Hence, from a perspective that assumes that time is a scarce resource, this finding may still imply that teams focusing on dysfunctional reflexive behaviors in revolutionary periods might fall behind those teams that showed the appropriate behaviors.

Testing the third hypothesis revealed somewhat mixed results. The hypothesis predicted a positive relation between quantitative reflection and performance during equilibrium periods. We did not find any of the proposed relations for T2 and T3, but for T5 and T6 (see models three and four). Based on our discussion regarding the first hypothesis and the question whether the T2 and T3 can be conceived as equilibrium periods, we rather argue in favor of this hypothesis. However, further research needs to investigate this finding and the related hypothesis.

The fourth hypothesis predicted a negative relation between qualitative reflection and performance during equilibrium periods. Our analysis only revealed one significant relation at T6. However, as this hypothesis was also based on equilibrium periods, the same restrictions we discussed regarding hypotheses 2 and 3 also apply to this hypothesis.

Conclusively, our findings indicate that there are particular reflexive behaviors supporting team adaptation in revolutionary and equilibrium periods, whereas others have no, or even a negative effect on performance. Based on the assumption that teams have only a limited amount of time, teams that focus on neutral or detrimental reflexive behaviors might therefore stay behind their potential, as they cannot invest their available time effectively, and may ultimately fall behind other teams.

4.8.1 Implications and Future Research

In spite of a considerable number of studies that investigated punctuated equilibrium theory, this study is among the first that used the theory's conceptualization of change to examine the relationship between team processes and outcomes. Moreover, as our results provide new insights into the previously found mixed effects of team reflection on performance, we argue that the application of the theories principles of contingencies on environmental demands may also help to examine mixed effects of processes on performance in other research areas (e.g., within team conflicts, de Wit, Greer, & Jehn, 2012). Furthermore, multiple implications can be drawn, which will be elaborated in the following sections.

4.8.1.1 Multi-Period, Multi-Theoretical.

First, we showed that the relation between the process of team reflection and performance is contingent on varying demands that emerge from punctuated equilibrium over multiple time-periods. Notably, as Humphrey and Aime (2014) argued in their extensive review on organizational behavior, the results underline the fundamental importance of studying adaptation in a multi-period, multi-theoretical setting, because cross-sectional and mono-theoretical research can only produce isolated snapshots, and thereby fail to capture the 'big picture'. For instance, based on the naïve assumption that team reflection should generally improve team performance, the present study would have reproduced the contradicting results that surfaced in previous team reflection research (cf. Konradt et al., 2016; Moreland & McMinn, 2010). We therefor again emphasize the mandatory need of longitudinal studies in organizational research.

4.8.1.2 Multi-Level.

However, Humphrey and Aime also argued, that organizational research should account for possible multi-level structures (cf. González-Romá & Hernández, 2017). Since we

used aggregated team measures, we assumed that all team members shared similar perceptions of the team reflexive process. Thus, we did not integrate the individual-level perspective that might pertain additional information (Humphrey & Aime, 2014). Moreover, as the obtained team agreement values were only judged as acceptable, there may have be substantial differences in the within-team perception of team reflexive processes. For instance, it may also be plausible to assume that teams with low agreement values are more effective. This notion is supported by previous reflection research, which has shown that the interaction of minority dissent and team reflection is related to team performance (De Dreu, 2002). On the other hand, teams with high agreement values might also be more effective in their reflection, because they assess their reflective processes more realistically, eventually resulting in smaller error estimations in statistical models. Future research should therefore investigate whether lower agreement-values in team reflection simply results from different perceptions in the form of random errors in our measurements, or whether systematical influences may surface, which are currently not captured by our models.

4.8.1.3 Deep Structures.

In this study, we build our hypothesis upon the concept of deep structures (Gersick, 1991, Tushman & Romanelli, 1985), that are either developed through exploration-related processes, or maintain by exploitation-related processes. However, in spite of the usefulness of the concept to deduce the hypothesis, we used the term from a rather descriptive perspective. Surprisingly, multiple results from our study provide evidence for the usefulness of deep structures beyond a solely descriptive concept. For instance, we found several lag-2 reflection-performance relations that significantly explained variance above the direct effects of reflection at the given time point (i.e., information evaluation at T4 significantly explained variance above information evaluation at T5). Consequently, previous team reflection somehow affects performance measured at the subsequent measurement point. We therefore

argue that deep structures might help explain this result. As quality and quantity of information evaluation at T4 probably enabled/hindered teams to developed better/worse solutions, this might have led to the establishment of a more/less functional deep structure that still influenced performance during later times. It might therefore be of great value to develop tools that allow a more detailed analysis of a team's (or organization's) deep structure, in order to provide functional advice and purposeful intervention strategies.

Moreover, another interesting finding in this context is the continuous decline in performance after the punctuation between T3 and T4. Although we did expect that the breaking of the equilibrium would result in a short decline in performance, the punctuations appears to have affected multiple teams in such a way that their performance continued to decline, which might be due to the initial formation of dysfunctional deep structures. Unfortunately, our sample size was too small to conduct further analysis using growth mixture modeling (Shiyko, Ram, & Grimm, 2012), to explore whether longitudinal performance development patterns (i.e., performance trajectories, Lindsley, Brass, & Thomas, 1995) can be identified in our data. In contrast to the analytical method we used in this study, which assumes that all teams develop equally over the examined measurement points, growth mixture modeling supports the identification of subgroups, that potentially develop on different performance trajectories. Slopes and intercepts of these trajectories can be conceived as latent variables, and hence, can be predicted by other focal variables, as for example team reflection. Further research might therefore identify typical patterns of adaptation or maladaptation, and help to develop a form of early warning system that supports team adaptation in dynamic environments.

An important follow up question is moreover, whether a team that initially did not show the right reflexive activities, may be able to catch up with more successful teams, by breaking equilibria and transforming the dysfunctional deep structure. Earlier research has

already provided evidence for the potential of leadership interventions to initiate revolutionary periods (Morgeson, 2005; Mudambi & Swift, 2014). Moreover, Håkonsson, Eskildsen, Argote, Mønster, Burton, and Obel (2015) reported that declines in performance increased the likelihood of teams exploring new routines, rather than exploiting existing ones. However, these questions have mainly been targeted from a management perspective.

4.8.1.4 Exploration and Exploitation Processes.

Hence, future research should seek to identify additional team processes that contribute to the development or maintenance of deep structures. West (2000), for instance, argued that the transition process of planning, and the action process of implementation, directly build upon preceding team reflection. However, both processes are rather scarcely represented in empirical organizational research (see for exceptions Fisher, 2014; Gevers, van Eerde, & Rutte, 2001). Similarly, Anderson, Potočnik, and Zhou (2014) reported in their review on innovation and creativity, which are closely connected to exploration and exploitation, that "notably few studies have examined within-team innovation processes as they unfold over time" (p. 1310). Moreover, in the more recent review on team innovation (van Knippenberg, 2017), team processes also rather appear as a side note. It might therefore be of great value to extend the ideas presented in this study and identify relevant team innovation processes (i.e., exploration and exploitation related), and understand their relations to outcomes.

4.8.1.5 Ambidexterity in Punctuated Equilibrium.

Finally, the question should be raised, whether a binary view on punctuated equilibrium, that is conceived of episodes that are either revolution, or equilibrium, is appropriate, as different degrees of revolution could persist after a punctuation. A declining revolutionary period might for instance imply some form of transition period between revolution and equilibrium, which is connected to different or constantly shifting

requirements. It is assumed in innovation (Anderson et al., 2014), that successful teams and organizations can balance explorative and exploitive activities. Similarly, a form of ambidexterity in reflexive behaviors, requiring the right degree of both qualitative and quantitative reflection at the same time, might help our understanding of the transition from revolution to equilibrium. For instance, teams could split up into smaller sub-groups, allowing the groups to focus on different forms of reflection and integrate the resulting insights more effectively. Hence, these teams might perform even better than teams focusing collectively on a single reflexive behavior. However, in their review, Gupta et al. (2006) already elaborated that building subsystems to achieve ambidexterity may only be viable when the subsystems are loosely coupled.

4.8.2 Limitations

There are three limitations to our findings that may have impacted the validity of our findings, and may therefore be targeted in future research. The first limitation concerns our sample size and the use of a significance level of .1. This might have led to false-positive identifications of several relations (increase of type I error rate). However, we thereby reduced the risk of missing relevant relations which were not significant because of our sample size. Future research should therefore investigate the robustness of our findings.

The second limitation results from the use of questionnaires as primary data source. Although the items are carefully framed as qualitative and quantitative items, the relatively large correlations between the quality and quantity dimensions indicate that the participants had difficulties to distinguish between both constructs. Moreover, questionnaires are only retrospective self-reports. Answers are therefore potentially biased (as for example because of social desirability) and may have led to distorted results (Podsakoff & Organ, 1986). Future research should therefore consider the use of more objective data sources that may complement the questionnaire-based data.

Thirdly, although the team reflexive processes of information seeking and information are unquestionably connected, we analyzed their effects on performance separately to reduce overall model complexity. Thus, future research should explore whether the effects of both processes are incremental (i.e., that information evaluation explains variance above information seeking), or whether more complex interaction patterns have to be assumed. However, previous reflection research (Gabelica, Van den Bossche, Segers, & Gijselaers, 2014) already found that teams rarely complete the relatively intuitive cycle of reflecting, planning, and implementing. Moreover, teams also tend to skip back and forth between these processes. Since information seeking and information evaluation behaviors are even more connected, their interplay might therefore be even more complex.

4.9. References

Aime, F., Humphrey, S., DeRue, D. S., & Paul, J. B. (2014). The riddle of heterarchy: Power transitions in cross-functional teams. *Academy of Management Journal, 57*, 327-352. doi:10.5465/amj.2011.0756

Anderson, N., Potočnik, K., & Zhou, J. (2 014). Innovation and creativity in organizations: A state-of-the-science review and prospective commentary. *Journal of Management, 40,* 1297-1333. doi:10.1177/0149206314527128

Arrow, H. (1997). Stability, bistability, and instability in small group influence patterns. *Journal of Personality and Social Psychology*, *72*(1), 75-85.

Baard, S. K., Rench, T. A., & Kozlowski, S. W. (2013). Performance adaptation: A theoretical integration and review. *Journal of Management, 40,* 48-99. doi:10.1177/0149206313488210

Bartunek, J. M., & Woodman, R. W. (2015). Beyond Lewin: Toward a temporal approximation of organization development and change. *Annual Review of Organizational Psychology and Organizational Behavior, 2,* 157-182. doi:10.1146/annurev-orgpsych-032414-111353

Brown, S. L., & Eisenhardt, K. M. (1997). The art of continuous change: Linking complexity theory and time-paced evolution in relentlessly shifting organizations. *Administrative Science Quarterly, 42,* 1-34. doi:10.2307/2393807

Burtscher, M. J., Manser, T., Kolbe, M., Grote, G., Grande, B., Spahn, D. R., & Wacker, J. (2011). Adaptation in anaesthesia team coordination in response to a simulated critical event and its relationship to clinical performance. *British Journal of Anaesthesia, 106,* 801-806. doi:10.1093/bja/aer039

Burtscher, M. J., Wacker, J., Grote, G., & Manser, T. (2010). Managing nonroutine events in anesthesia: The role of adaptive coordination. *Human Factors: The Journal of the*

Human Factors and Ergonomics Society, 52, 282-294. doi:10.1177/0018720809359178

Carver, C.S., & Scheier, M.F. (1998). *On the Self-Regulation of Behavior*. Cambridge, England: Cambridge University Press.

Chang, A., Bordia, P., & Duck, J. (2003). Punctuated equilibrium and linear progression: Toward a new understanding of group development. *Academy of Management Journal, 46*, 106-117. doi:10.2307/30040680

De Dreu, C. K. W. (2002). Team innovation and team effectiveness: The importance of minority dissent and reflexivity. *European Journal of Work and Organizational Psychology, 11,* 285-298. doi:10.1080/13594320244000175

De Wit, F. R., Greer, L. L., & Jehn, K. A. (2012). The paradox of intragroup conflict: a meta-analysis. *Journal of Applied Psychology, 97*, 360-390. doi:10.1037/a0024844

DeShon, R.P., Kozlowski, S.W. J., Schmidt, A.M., Milner, K.R., & Wiechmann, D. (2004). A multiple goal, multilevel model of feedback effects on the regulation of individual and team performance. *Journal of Applied Psychology, 89,* 1035–1055. doi:10.1037/0021-9010.89.6.1035

Fisher, D. M. (2014). Distinguishing between taskwork and teamwork planning in teams: Relations with coordination and interpersonal processes. *Journal of Applied Psychology, 99*, 423-436. doi:10.1037/a0034625

Gabelica, C., Van den Bossche, P., Segers, M., & Gijselaers, W. (2014). Dynamics of team reflexivity after feedback. *Frontline Learning Research, 2*, 64–91. doi:10.14786/flr.v2i2.79

Gersick, C. J. (1989). Marking time: Predictable transitions in task groups. *Academy of Management Journal*, *32*, 274-309. doi:10.2307/256363

Gersick, C. J. (1991). Revolutionary change theories: A multilevel exploration of the

punctuated equilibrium paradigm. *Academy of Management Review*, *16*, 10-36. doi:10.5465/AMR.1991.4278988

Gevers, J. M., van Eerde, W., & Rutte, C. G. (2001). Time pressure, potency, and progress in project groups. *European Journal of Work and Organizational Psychology*, *10*, 205-221. doi:10.1080/13594320143000636

Gupta, A. K., Smith, K. G., & Shalley, C. E. (2006). The interplay between exploration and exploitation. *Academy of Management Journal*, *49*, 693-706. doi:10.5465/AMJ.2006.22083026

Gurtner, A., Tschan, F., Semmer, N. K., & Nägele, C. (2007). Getting groups to develop good strategies: Effects of reflexivity interventions on team process, team performance, and shared mental models. *Organizational Behavior and Human Decision Processes*, *102*, 127-142. doi:10.1016/j.obhdp.2006.05.002

Håkonsson, D., Eskildsen, J. K., Argote, L., Mønster, D., Burton, R. M., & Obel, B. (2015). Exploration versus exploitation: Emotions and performance as antecedents and consequences of team decisions. *Strategic Management Journal*, *37*, 985-1001. doi:10.1002/smj.2380

Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods*, *1*, 104-121. doi:10.1177/109442819800100106

Hollenbeck, J. R., Moon, H., Ellis, A. P., West, B. J., Ilgen, D. R., Sheppard, L., ... & Wagner III, J. A. (2002). Structural contingency theory and individual differences: examination of external and internal person-team fit. *Journal of Applied Psychology*, *87*, 599-606. doi:10.1037/0021-9010.87.3.599

Humphrey, S. E., & Aime, F. (2014). Team microdynamics: Toward an organizing approach to teamwork. *The Academy of Management Annals*, *8*, 443–503.

doi:10.1080/19416520.2014.904140

Ilgen, D. R., Hollenbeck, J. R., Johnson, M., & Jundt, D. (2005). Teams in organizations: from input-process-output models to IMOI models. *Annual Review of Psychology, 56*, 517–543. doi:10.1146/annurev.psych.56.091103.070250

Kaplan, D., & Depaoli, S. (2012). Bayesian structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of Structural Equation Modeling* (pp. 650-673). New York: The Guildford Press.

Konradt, U., & Eckardt, G. (2016). Short-term and long-term relations among reflection and performance in teams: Evidence from a four-wave longitudinal study. *European Journal of Work and Organizational Psychology, 25*, 804–818. doi:10.1080/1359432X.2016.1160058

Konradt, U., Otte, K.-P., Schippers, M. C., & Steenfatt, C. (2017). Reflexivity in teams – A review and new perspectives. *The Journal of Psychology: Interdisciplinary and Applied, 150*, 153–174. doi:10.1080/00223980.2015.1050977

Lang, J. W., & Bliese, P. D. (2009). General mental ability and two types of adaptation to unforeseen change: applying discontinuous growth models to the task-change paradigm. *Journal of Applied Psychology, 94*, 411-428. doi:10.1037/a0013803

Lawrence, P. R., & Lorsch, J. W. (1967). Differentiation and integration in complex organizations. *Administrative Science Quarterly, 12*(1), 1-47.

LePine, J. A. (2003). Team adaptation and postchange performance: Effects of team composition in terms of members' cognitive ability and personality. *Journal of Applied Psychology, 88*, 27-39. doi:10.1037/0021-9010.88.1.27

LePine, J. A. (2005). Adaptation of teams in response to unforeseen change: effects of goal difficulty and team composition in terms of cognitive ability and goal orientation. *Journal of Applied Psychology, 90*, 1153-1167. doi:10.1037/0021-9010.90.6.1153

LePine, J. A., Colquitt, J. A., & Erez, A. (2000). Adaptability to changing task contexts: Effects of general cognitive ability, conscientiousness, and openness to experience. *Personnel Psychology, 53*, 563-593. doi:10.1111/j.1744-6570.2000.tb00214.x

LePine, J. A., Piccolo, R. F., Jackson, C. L., Mathieu, J. E., & Saul, J. R. (2008). A meta-analysis of teamwork processes: tests of a multidimensional model and relationships with team effectiveness criteria. *Personnel Psychology, 61*, 273-307. doi:10.1111/j.1744-6570.2008.00114.x

Li, Y., Vanhaverbeke, W., & Schoenmakers, W. (2008). Exploration and exploitation in innovation: Reframing the interpretation. *Creativity and Innovation Management, 17*, 107-126. doi:10.1111/j.1467-8691.2008.00477.x

Lim, S. G. S., & Murnighan, J. K. (1994). Phases, deadlines, and the bargaining process. *Organizational Behavior and Human Decision Processes*, *58*, 153-171. doi:10.1006/obhd.1994.1032

Lindell, M. K., Brandt, C. J., & Whitney, D. J. (1999). A revised index of interrater agreement for multi-item ratings of a single target. *Applied Psychological Measurement, 23*, 127–135. doi:10.1177/01466219922031257

Lindsley, D. H., Brass, D. J., & Thomas, J. B. (1995). Efficacy-performing spirals: A multilevel perspective. *Academy of Management Review, 20*, 645-678. doi:10.5465/AMR.1995.9508080333

March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science, 2*, 71-87. doi:10.1287/orsc.2.1.71

Marks, M. A., Mathieu, J. E., & Zaccaro, S. J. (2001). A temporally based framework and taxonomy of team processes. *Academy of Management Review, 26*, 356–376. doi:10.5465/AMR.2001.4845785

Mathieu, J. E., Tannenbaum, S. I., Donsbach, J. S., & Alliger, G. M. (2014). A review and

integration of team composition models moving toward a dynamic and temporal framework. *Journal of Management, 40*, 130-160. doi:10.1177/0149206313503014

Mathieu, J., Maynard, M. T., Rapp, T., & Gilson, L. (2008). Team effectiveness 1997-2007: A review of recent advancements and a glimpse into the future. *Journal of Management, 34*, 410-476. doi:10.1177/0149206308316061

Maynard, M. T., Kennedy, D. M., & Sommer, S. A. (2015). Team adaptation: A fifteen-year synthesis (1998–2013) and framework for how this literature needs to "adapt" going forward. *European Journal of Work and Organizational Psychology*, *24*, 652-677. doi:10.1080/1359432X.2014.1001376

Morgeson, F. P. (2005). The external leadership of self-managing teams: intervening in the context of novel and disruptive events. *Journal of Applied Psychology, 90*, 497-508. doi:10.1037/0021-9010.90.3.497

Morgeson, F. P., Mitchell, T. R., & Liu, D. (2015). Event system theory: An event-oriented approach to the organizational sciences. *Academy of Management Review, 40*, 515-537. doi:10.5465/amr.2012.0099

Moreland, R. L., & McMinn, J. G. (2010). Group reflexivity and performance. *Advances in Group Processes, 27*, 63–95. doi:10.1108/s0882-6145(2010)0000027006

Mudambi, R., & Swift, T. (2014). Knowing when to leap: Transitioning between exploitative and explorative R&D. *Strategic Management Journal, 35*, 126-145. doi:10.1002/smj.2097

Muthén, B. & Asparouhov, T. (2012). Bayesian SEM: A more flexible representation of substantive theory. *Psychological Methods, 17*, 313-335. doi:10.1037/a0026802

Muthén, L. K., & Muthén, B. O. (2017). *Mplus User's Guide. Seventh Edition*. Los Angeles, CA: Muthén & Muthén.

Okhuysen, G. A., & Waller, M. J. (2002). Focusing on midpoint transitions: An analysis of

boundary conditions. *Academy of Management Journal*, *45*, 1056-1065. doi:10.2307/3069330

Otte, K.-P., Konradt, U., Garbers, Y., & Schippers, M. C. (2017). Development and validation of the REMINT: a reflection measure for individuals and teams. *European Journal of Work and Organizational Psychology*, *26*, 299-313. doi:10.1080/1359432X.2016.1261826

Podsakoff, P. M., & Organ, D. W. (1986). Self-reports in organizational research: Problems and prospects. *Journal of Management*, *12*, 531-544. doi:10.1177/014920638601200408

Pulakos, E. D., Arad, S., Donovan, M. A., & Plamondon, K. E. (2000). Adaptability in the workplace: development of a taxonomy of adaptive performance. *Journal of Applied Psychology*, *85*, 612-624. doi:10.1037//0021-9010.85.4.612

Randall, K. R., Resick, C. J., & DeChurch, L. A. (2011). Building team adaptive capacity: the roles of sensegiving and team composition. *Journal of Applied Psychology*, *96*, 525-540. doi:10.1037/a0022622

Romanelli, E., & Tushman, M. L. (1994). Organizational transformation as punctuated equilibrium: An empirical test. *Academy of Management Journal*, *37*, 1141-1166. doi:10.2307/256669

Sastry, M. A. (1997). Problems and paradoxes in a model of punctuated organizational change. *Administrative Science Quarterly*, *42*, 237-275. doi:10.2307/2393920

Savelsbergh, C., Van der Heijden, B. I. J. M., & Poell. R. F. (2009). The development and empirical validation of a multi-dimensional measurement instrument for team learning behaviors. *Small Group Research*, 40, 578–607. doi:10.1177/1046496409340055

Schippers, M. C., Homan, A. C., & van Knippenberg, D. (2013). To reflect or not to reflect: Prior team performance as a boundary condition of the effects of reflexivity on

learning and final team performance. *Journal of Organizational Behavior, 34*, 6–23. doi:10.1002/job.1784

Schippers, M. C., West, M. A., & Dawson, J. F. (2015). Team reflexivity and innovation: The moderating role of team context. *Journal of Management*, 41, 769–788. doi:10.1177/0149206312441210

Seers, A., & Woodruff, S. (1997). Temporal pacing in task forces: Group development or deadline pressure? *Journal of Management, 23*, 169-187. doi:10.1016/S0149-2063(97)90042-8

Shiyko, M. P., Ram, N., & Grimm, K. J. (2012). An overview of growth mixture modeling: A simple nonlinear application in OpenMx. In R. H. Hoyle (Ed.), *Handbook of Structural Equation Modeling* (pp. 532-546). New York: Guilford Press.

Song, X. Y., & Lee, S. Y. (2012). *Basic and Advanced Bayesian Structural Equation Modeling: With Applications in the Medical and Behavioral Sciences*. London: Wiley. doi:10.1002/9781118358887

Stachowski, A. A., Kaplan, S. A., & Waller, M. J. (2009). The benefits of flexible team interaction during crises. *Journal of Applied Psychology, 94*, 1536-1543. doi:10.1037/a0016903

Suddaby, R., & Foster, W.M. (2017). History and Organizational Change. *Journal of Management, 43*, 19-38. doi:10.1177/0149206316675031

Tilcsik, A., & Marquis, C. (2013). Punctuated generosity: How mega-events and natural disasters affect corporate philanthropy in US communities. *Administrative Science Quarterly, 58*, 111-148. doi:10.1177/0001839213475800

Tuckman, B. W., & Jensen, M. A. C. (1977). Stages of small-group development revisited. *Group & Organization Studies*, *2*(4), 419-427.

Tushman, M. L., & O'Reilly, C. A. (1996). The ambidextrous organizations: Managing

evolutionary and revolutionary change. *California Management Review*, *38*(4), 8-30.

Tushman, M. L., & Romanelli, E. (1985). Organizational evolution: A metamorphosis model of convergence and reorientation. *Research in Organizational Behavior, 7,* 171-222.

Tushman, M. L., & Rosenkopf, L. (1992). Organizational determinants of technological-change-toward a sociology of technological evolution. *Research in Organizational Behavior*, *14*, 311-347.

van Knippenberg, D. (2017). Team innovation. *Annual Review of Organizational Psychology and Organizational Behavior*, *4*, 211-233. doi:10.1146/annurev-orgpsych-032516-113240

Waller, M. J. (1999). The timing of adaptive group responses to nonroutine events. *Academy of Management Journal*, *42*, 127-137. doi:10.2307/257088

Waller, M. J., Gupta, N., & Giambatista, R. C. (2004). Effects of adaptive behaviors and shared mental models on control crew performance. *Management Science*, *50*, 1534-1544. doi:10.1287/mnsc.1040.0210

Waller, M. J., Zellmer-Bruhn, M. E., & Giambatista, R. C. (2002). Watching the clock: Group pacing behavior under dynamic deadlines. *Academy of Management Journal*, *45*, 1046-1055. doi:10.2307/3069329

Weick, K. E., & Quinn, R. E. (1999). Organizational change and development. *Annual Review of Psychology*, *50*, 361-386. doi:10.1146/annurev.psych.50.1.361

West, M. A. (2000). Reflexivity, revolution, and innovation in work teams. In M. M. Beyerlein, D. Johnson, & S. T. Beyerlein (Eds.), *Product Development Teams* (Vol. 5, pp. 1–29). Stamford, Connecticut: JAI Press.

Wollin, A. (1999). Punctuated equilibrium: Reconciling theory of revolutionary and incremental change. *Systems Research and Behavioral Science*, *16*, 359-367. doi:10.1002/(SICI)1099-1743(199907/08)16:4<359::AID-SRES253>3.0.CO;2-V

Zellmer-Bruhn, M. E. (2003). Interruptive events and team knowledge acquisition. *Management Science, 49,* 514-528. doi:10.1287/mnsc.49.4.514.14423

Zajac, S., Gregory, M. E., Bedwell, W. L., Kramer, W. S., & Salas, E. (2014). The cognitive underpinnings of adaptive team performance in ill-defined task situations: A closer look at team cognition. *Organizational Psychology Review, 4,* 49-73. doi:10.1177/2041386613492787

Zyphur, M. J. & Oswald, F. L. (2015). Bayesian estimation and inference: A user's guide. *Journal of Management, 41,* 390–420. doi:10.1177/0149206313501200

Kapitel 5

Abschließende Diskussion

5.1 Abschließende Diskussion

Zentrales Ziel dieser Dissertation war die theoretische und empirische Identifizierung möglicher Ursachen für die heterogene Befundlage hinsichtlich der Beziehung zwischen Team Reflexion und Leistung. Als erstes wurde dafür der Omnibusterm *Reflexivität* in Form eines Frameworks neu gefasst. Der Vorteil eines Frameworks ist dabei die Integration verschiedener reflexiver Prozesse in einem gemeinsamen Rahmen, ohne dabei die differenziertere Betrachtung und Erfassung der einzelnen Komponenten zu verlieren. Auf den Reflexionsprozess fokussierend, wurde daraufhin die neue, fragebogenbasierte Skala REMINT (Reflection Measure for Individuals and Teams) entwickelt und ausführlich anhand mehrerer Stichproben aus verschiedenen Kontexten validiert. Die Skala bietet dabei im Vergleich zu vorherigen Verfahren eine präzisere und gleichzeitig ökonomische Erfassung des Reflexionsprozesses, und erlaubt die wichtige Differenzierung zwischen den Dimensionen Informationssuche und Informationsentwicklung, sowie deren Qualität und Quantität. Auf Basis dieser Differenzierung konnten in den beiden folgenden Studien wichtige Hinweise dazu gefunden werden, dass die bisherige, unidimensionale Betrachtungsweise des Reflexionsprozesses ungeeignet war, um den Reflexions-Leistungszusammenhang systematisch zu analysieren. Es scheint daher unabdingbar zu sein, mehrere Facetten des Reflexionsprozesses gleichzeitig zu betrachten.

Darüber hinaus konnte gezeigt werden, dass die Beziehung zu Leistung nicht nur von der Qualitäts-Quantitätsdifferenzierung abhängig ist, sondern auch von kontextuellen Drittvariablen, wie beispielsweise vorangehender Leistung, verfügbarer Zeit, oder mit Umweltveränderungen einhergehenden, wechselnden Anforderungen. Aufbauend auf diesen Befunden lassen sich wichtige theoretische und praktische Implikationen ableiten.

5.2 Theoretische Implikationen

Ein zentrales Ergebnis dieser Dissertation ist die erfolgreiche Differenzierung zwischen der Qualität und der Quantität von Team Reflexion. Durch die gleichzeitige Berücksichtigung beider Dimensionen konnten in Studie 2 und Studie 3 Bedingungen identifiziert werden, unter denen bestimmte reflexive Prozesse einen positiven, keinen, oder sogar einen negativen Einfluss auf die Leistung eines Teams haben können. Aufbauend auf diesem Ergebnis könnte es daher sinnvoll sein, zur Erfassung der reflexive Teamprozesse *Planung* und *Implementation* eine ähnliche Vorgehensweise zu wählen.

In der zweiten Studie konnte beispielsweise beobachtet werden, dass Teams einen großen Teil ihrer Zeit auf die Planung von Strategien verwenden können. Solange diese Strategien allerdings nicht ausreichend detailliert geplant sind, könnten im weiteren Verlauf der Bearbeitung einer Aufgabe schnell Schwierigkeiten auftreten. Bisher hat sich die organisationale Planungsliteratur konzeptionell eher mit der Aufteilung von Planung in aufgabenbezogene und teambezogene (Fisher, 2014), oder bewusste, eventuelle und reaktive Planung (DeChurch & Haas, 2008; Marks, Mathieu, & Zaccaro, 2001) auseinandergesetzt. DeChurch und Haas (2008) berichteten beispielsweise, dass vor einer Aufgabe stattfindende Planungsprozesse im Team längst nicht so positive Effekte auf die Leistung hatten, wie reaktive Adjustierungen (reaktives planen). Aktuelle Untersuchungen zu viel diskutierten deutschen Großprojekten, wie der Elbphilharmonie oder dem Berlin Brandenburg Flughafen (Kostka & Fiedler, 2016), weisen allerdings deutlich darauf hin, dass reaktive Planung sehr schnell zu unerwünschten Folgen führen kann und vor allem bei komplexen Aufgaben nur in einem begrenzten Rahmen sinnvoll ist. Daher könnte die Aufteilung in qualitative und quantitative Planungselemente die in DeChurch & Haas gefundene Unterlegenheit von vorausgehenden Planungsprozessen möglicherweise erklären (siehe auch Smith & Locke, 1990).

Hinsichtlich des Implementationsprozesses sind zweierlei Punkte anzusprechen. Erstens hat bisherige Forschung zur Implementation in der Reflexivitätsliteratur nicht den tatsächlichen Prozess der Implementation erfasst (siehe auch Studie 2). Stattdessen wurde darauf fokussiert, inwiefern Teams die im (nicht betrachteten) Planungsprozess beschlossenen Handlungen (z.B. Strategien) umgesetzt haben. Auch wenn diese Vorgehensweise funktionell ist und bisher vergleichsweise zuverlässig funktioniert hat (siehe auch Konradt, Schippers, Garbers, & Steenfatt, 2015), sollte sich zukünftige Forschung mit alternativen Messmethoden auseinandersetzen, um eine validiere Erfassung des tatsächlichen Implementierungsprozesses zu ermöglichen. Eine Alternative bestünde beispielsweise in der Nutzung objektiver Datenquellen (siehe Gurtner, Tschan, Semmer, & Nägele, 2007), wie sie im in der dritten Studie verwendeten Unternehmensplanspiel TopSim, in Form von Unternehmenszahlen (z.B. Preise, Fertigungskapazitäten, Forschung- und Entwicklungsinvestitionen) zur Verfügung stehen.

Zweitens wäre daher auch bei der Messung von Implementierung die Konzeptualisierung einer Quantitäts- und Qualitätsdimension denkbar. Quantitative Implementation entspräche dabei zum Beispiel der Anzahl an Änderungen, die ein Team im Vergleich zur vorangegangenen Performance Episode vorgenommen hat, wohingegen qualitative Implementation die Tiefe bzw. Ausführlichkeit einzelner Änderungen erfassen würde. Generell lässt sich darüber hinaus an dieser Stelle festhalten, dass der häufige Fokus auf quantitative Elemente im organisationalen Handeln und Erleben möglicherweise nur einen Teil der tatsächlichen Zusammenhänge aufdeckt. Die zusätzliche Betrachtung qualitativer Elemente könnte daher auch außerhalb der Reflexivitätsforschung weitere Varianz von Kriterien aufklären, oder sogar zur Identifikation von bisher noch nicht betrachteten Effekten führen.

Die zweite relevante Erkenntnis dieser Dissertation ist der fundamentale Einfluss der zeitlichen Dimension auf die Zusammenhänge zwischen Reflexion und Leistung. Betrachtet man beispielsweise die Ergebnisse der dritten Studie, so hätte jeder der sechs Messzeitpunkte für sich genommen ein sehr schwer interpretierbares Bild ergeben. Erst durch die Berücksichtigung der Zeit in Form von früheren Performance Episoden (Marks et al., 2001) und systematischen Umweltveränderungen (Gersick, 1991) konnten Muster hinter diesen Zusammenhängen aufgezeigt werden. Es lässt sich daher festhalten, dass das Einbeziehen der zeitlichen Dimension unabdingbar ist für eine valide Beschreibung und Erklärung von Zusammenhängen im organisationalen Kontext.

Auch wenn dabei nicht zwangsweigerlich auf die Punctuated Equilibrium Theorie zurückgegriffen werden muss, hat der Einsatz der Theorie in der dritten Studie aufgrund der grundlegenden Konzeptualisierung vom Wandel, in Form von Equilibriums- und Revolutionsperioden, zum Verständnis von zuvor nur schwer erklärbaren Zusammenhängen beigetragen. Da sich die heutige, globale Arbeitswelt aufgrund eines sich immer noch weiter beschleunigenden technischen Wandels und multipler, globaler Risikofaktoren (Wagner & Disparte, 2016) auch zukünftig durch tiefgreifende Erschütterungen (Punctuationen) auszeichnen wird, könnte die Theorie auch für die weitere Forschung sinnvolle Orientierungspunkte bieten.

Die dritte Erkenntnis und damit einhergehende Implikation, die an dieser Stelle festzuhalten ist, geht aus der Fokussierung auf Team Prozesse hervor. Im Gegensatz zu einem Großteil der Forschung zu selbst-regulatorischen Anpassung von Teams hat sich diese Arbeit nicht auf Antezedenzien (z.B. mentale Leistungsfähigkeit, Bliese & Lang, 2009) von Anpassungsfähigkeit konzentriert, sondern auf zugrundeliegenden Mechanismen (cf. Baard, Rench, & Kozlowski, 2014). Dabei zeigen die beispielsweise in Studie zwei berichteten Ergebnisse, dass Team Reflexion und Implementierung über vorangehende Leistung hinaus

30% der Kriteriumsvarianz aufklären konnten. Dieses Resultat verdeutlicht daher die Relevanz dieser oftmals vernachlässigten (Maynard, Kennedy, & Sommer, 2015) Prozess-Perspektive. In Anlehnung an Arbeiten zur Strukturierung von Teamprozessen (bspw. Marks et al., 2001; Mathieu, Maynard, Rapp, & Gilson, 2008; DeChurch & Mesmer-Magnus, 2010), oder auch der in der dritten Studie vorgeschlagenen Differenzierung zwischen Explorations- und Exploitationsprozessen, könnte eine stärker prozessorientierte Forschung für die Organisationsliteratur daher einen Gewinn bedeuten.

5.3 Praktische Implikationen

Ergänzend lassen sich aus den Ergebnissen dieser Dissertation auch praktisch relevante Implikationen ziehen. Tannenbaum und Cerasoli argumentierten 2013 auf Basis meta-analytischer Befunde, dass Reflexionstechniken (sogenannte after-action-reviews) die individuelle Leistung von Mitarbeitern innerhalb einer Organisation um 20-25 Prozent steigern können. Die in der Meta-Analyse verwendeten Studien basierten allerdings fast ausschließlich auf laborbasierten, hochstrukturierten Untersuchungen, die Reflexion in der Form Vorhanden/Nicht vorhanden manipuliert haben. Darüber hinaus enthielt diese Meta-Analyse keine einzige Untersuchung aus dem Reflexivitätskontext und verfolgte ebenso den bereits mehrfach erwähnten ‚One-size-fits-all' Ansatz.

In dieser Dissertation konnten wichtige, und praktisch höchst relevante Hinweise darauf gefunden werden, dass *selbst-gesteuerte* Team Reflexion einen positiven Effekt auf die Leistung haben kann. Für die angewandte Praxis (beispielsweise Führungstechniken oder Team-Trainingsmaßnahmen) sollte dennoch hervorgehoben werden, dass dieses Ergebnis als nicht für jede Situation gültig zu betrachten ist. Stattdessen ist es wichtig, sich der Einschränkungen bewusst zu sein, die aus einer dynamischen Umwelt resultieren. In dieser *können* sich bestimmte, reflexive Verhaltensweise zu einem gegebenen Zeitpunkt positiv auswirken. Zu einem späteren Zeitpunkt kann sich dieser Effekt allerdings bereits wieder

umkehren. An dieser Stelle muss ebenfalls betont werden, dass auch das Fehlen einer Relation zwischen Team Reflexion und Leistung relevant sein kann. Aufbauend auf der in Studie 2 präsentierten Annahme, dass Team Reflexion ein zeitintensiver Prozess ist, kann ein oberflächlich betrachtet nicht schadhafter Reflexionsprozess langfristig sich dennoch negativ auswirken, weil zeitliche Ressourcen falsch eingesetzt werden.

Als zweiten wichtigen Punkt für die Praxis ist hervorzuheben, dass in der zweiten Studie der Begriff qualitative Team Reflexion für die Praxis greifbarer gemacht werden konnte. Aufbauend auf dem Kodierschema wurden wichtige Hinweise präsentiert, dass die positiven Effekte von qualitativer und detaillierter Reflexion, von der tiefgehenden Evaluation von vorangehenden Strategien, Vorgehensweisen, und Taktiken ausgehen. Auch wenn an dieser Stelle noch weitere Überprüfungen in anderen Kontexten und Aufgabentypen nötig sind, bietet dieser Befund Ansatzpunkte für die Konzeption von Trainings und Interventionsmaßnahmen.

5.4 Limitationen und zukünftige Forschung

Trotz des Umfangs der bisherigen Untersuchungen, konnte in dieser Dissertation aus dem Reflexivitätsframework (siehe Studie 1) primär nur die Team Reflexion, in Form der Dimensionen Qualität und Quantität von Team Informationssuche und Team Informationsevaluation abgedeckt werden. Folgende wichtige Fragen sollten daher in zukünftiger Forschung bearbeitet werden.

Zuerst ist daher zu erwähnen, dass der Planungsprozess im Reflexivitätsframework in dieser Arbeit nicht berücksichtigen werden konnte, ohne den Umfang massiv zu steigern. Auch wenn der Planungsprozess in der bisherigen Reflexivitätsforschung mit einer Ausnahme (Gevers, van Eerde, & Rutte, 2001) gar nicht berücksichtigt wurde, argumentieren viele Autoren (bspw. Marks et al., 2001) für die Relevanz des Prozesses. Eine systematische

Auseinandersetzung mit selbiger würde daher sicherlich einen großen Mehrwert für die Forschung bedeuteten.

Als zweite Limitation ist anzumerken, dass die gewählte, zeitliche Auflösung der Messungen (das heißt der Abstand zwischen zwei Messzeitpunkten) zwar über vieles in der bisherigen Reflexivitätsforschung hinausgeht, aber womöglich immer noch nicht hoch genug ist. Insbesondere wird der verwendete zeitliche Rhythmus durch den Einsatz der Performance Episoden vorgegeben (Marks et al., 2001). Da die Messungen allerdings immer am Ende einer Performance Episode vorgenommen wurden, sind möglicherweise relevante Elemente innerhalb einer Episode nicht erfasst worden, da es sich lediglich um eine zeitliche Aggregation über die gesamte Episode gehandelt hat. Ein verbreitetes Phänomen in der zweiten Studie war beispielsweise, dass sich Teams nach dem ersten Durchgang euphorisch in die Planung des zweiten Durchgangs gestürzt haben. Erst zum Ende der Reflexionsphase stellten manche schließlich fest, dass sie elementare Aspekte der Aufgabe noch nicht verstanden haben, weshalb sie verspätet begannen zu reflektieren. Es könnte daher sinnvoll sein von der Annahme auszugehen, dass sich Planung positiver auf ein Kriterium wie Leistung auswirkt (oder überhaupt erst), wenn das Team zuvor auf korrekte Art und Weise reflektiert. Dementsprechend könnte im besprochenen Beispiel ein Teil des Planungsprozesses nutzlos gewesen sein, und die Planung somit ineffektiv. Daher könnten Teams, die zu Beginn der Reflexionsphase reflektiert haben, erfolgreicher gewesen sein als Teams, die dies erst am Ende getan haben, was von der derzeitigen Konzeption und Messung nicht abgedeckt wurde.

Hieraus ergibt sich für zukünftige Forschung nicht nur die Frage nach der Relevanz des zeitlichen Auftretens eines Team Prozesses, sondern auch nach Interaktionen mit anderen Prozessen. Der Begriff *Interaktion* bezeichnet dabei das Zusammenwirken zweier oder mehr Variablen und ihr gemeinsamer Einfluss auf ein Kriterium. Wie aus der weit verbreiteten

Nutzung von hierarchischen Regressionsanalysen ersichtlich wird (z.B. De Dreu, 2007; Schippers, den Hartog, & van Knippenberg, 2013; Schippers, West, & Dawson, 2015), beruht ein Großteil der Annahmen in der Reflexivitätsforschung auf vergleichsweise einfachen, inkrementellen Modellen. Das heißt, es wird konträr zur Intuition sehr oft davon ausgegangen, dass sich die Prädiktoren innerhalb eines Modells nur zu einem zu vernachlässigenden Maß gegenseitig beeinflussen. Insgesamt stellt sich an dieser Stelle also die Frage, ob die Effekte mehrerer Prozesse rein additiv sind (d.h. mehrere Prädiktoren klären übereinander Varianz auf), oder ob sich aufgrund von zeitlicher Reihenfolge und Interaktion zwischen den Prozessen Effekte ergeben, die über die Additivität hinausgehen. Kozlowski und Klein (2000) differenzieren hierbei zwischen *Compositionseffekten* (rein additiv) und *Compilationseffekte* (über die Addition hinaus, emergent; siehe auch Fulmer & Ostroff, 2015). Zukünftige Forschung könnte daher mittels Sequenzanalysen (Gabadinho, Ritschard, Müller, & Studer, 2011) verschiedene Verhaltenssequenzen identifizieren. Diese Sequenzen könnten daher, im Gegensatz zu einer Aggregation über die gesamte Performance Episode, eine differenziertere Betrachtung der Effekte verschiedener Prozesse auf die Teamleistung erlauben.

Wie in Studie 3 allerdings deutlich ersichtlich wurde, können bereits vergleichsweise trivial anmutende Zusammenhänge zwischen zwei Variablen (bspw. Reflexion und Leistung) eine beeindruckende Komplexität entfalten, wenn man sie im Kontext einer dynamischen Umwelt betrachtet. Durch das hinzunehmen weiterer Prozesse können daher schnell die Grenzen der derzeitig verwendeten, fragebogenbasierten Messmethoden überschritten werden. Potenzielle Lösungsansätze bieten hier beispielsweise Simulationsstudien (Vancouver & Weinhardt, 2012), die bisher vergleichsweise selten Anwendung in der Organisationsforschung fanden, aber eine detailliertere Konzeption und Überprüfung von dynamischen Zusammenhängen erlauben.

Als dritte Limitation sind die vergleichsweise hohen Korrelationen zwischen den Qualitäts- und Quantitätsdimensionen zu erwähnen, die möglicherweise zur Verdeckung relevanter Zusammenhänge geführt haben könnten (vgl. Cole, Bedeian, Hirschfeld, & Vogel). Zukünftige Forschung sollte sich daher mit Methoden auseinandersetzen, wie die Qualitäts- und Quantitätsmaße in Fragebogenverfahren noch besser differenziert werden können.

Als vierte Limitation ist anzumerken, dass Studie 2 und Studie 3 auf vergleichsweise kleinen Stichproben basierten. Dadurch entstehende Nachteile konnten zwar teilweise durch die Verwendung des Bayes-Schätzers ausgeglichen werden, es muss aber dennoch notiert werden, dass die Ergebnisse weiterer Validierung bedürfen. In diesem Kontext ist auch die Kritik zu äußern, dass die Analysemethoden alle Teams immer noch vergleichsweise ähnlich behandelt haben. Durch dieses Vorgehen könnten möglicherweise Subpopulationen in den Stichproben verdeckt worden sein, innerhalb derer sich die Abläufe und Zusammenhänge fundamental unterscheiden. Mittels Growth Mixture Modeling (Shiyko, Ram, & Grimm, 2012) könnten diese Subpopulationen entdeckt, und separat analysiert werden. Allerdings steigen mit der Anzahl an Subpopulationen (d.h. Gruppen) die Anforderungen an die Stichprobengröße, wodurch die Anwendung dieser Analysemethode sich vor dem Hintergrund der Stichprobengrößen dieser Dissertation als schwierig erwiesen hätte.

Zuletzt fallen vor allem in Studie 2, aber auch in Studie 3 vergleichsweise geringe Übereinstimmungswerte bezüglich Team Reflexion innerhalb der Gruppen auf. Diese Werte können sich auf die Ergebnisse ausgewirkt haben, wobei die entstandene Fehlervarianz potenziell eher zu einer Unter- als Überschätzung der Zusammenhänge geführt hat. Mögliche Ursachen hierfür wurden in beiden Studien bereits diskutiert, weshalb an dieser Stelle nur darauf eingegangen werden soll, dass zukünftige Forschung nicht nur eine höhere zeitliche Auflösung anstreben sollte, sondern parallel auch die Prozesse und Interaktionen innerhalb der einzelnen Teams berücksichtigen muss (vgl. Chan, 1998; Humphrey & Aime, 2014).

5.5 Abschließende Bemerkung

Insgesamt wurden in der vorliegenden Dissertation eindeutige Hinweise gesammelt, dass der Zusammenhang zwischen Team Reflexivität und Leistung, und insbesondere Team Reflexion und Leistung, eine deutlich ausgeprägtere Komplexität aufweist, als bisher oftmals angenommen und dargestellt wurde. Die Ergebnisse bieten daher zusammen mit den neu entwickelten Messmethoden und erweiterten theoretische Modellen wichtige Ansatzpunkte für die zukünftige Forschung, um das Verständnis dieses relevanten Team Forschungsfeldes weiter voranzutreiben.

5.6 Quellen

Baard, S. K., Rench, T. A., & Kozlowski, S. W. (2013). Performance adaptation: A theoretical integration and review. *Journal of Management, 40*, 48-99. doi:10.1177/0149206313488210

Chan, D. (1998). Functional relations among constructs in the same content domain at different levels of analysis: A typology of composition models. *Journal of Applied Psychology, 83*, 234-246. doi:10.1037/0021-9010.83.2.234

Cole, M. S., Bedeian, A. G., Hirschfeld, R.R., & Vogel, B. (2011). Dispersion-composition models in multilevel research: A data-analytic framework. *Organizational Research Methods, 14*, 718-734. doi:10.1177/109442811038907

Cronin, M. A., Gonzalez, C., & Sterman, J. D. (2009). Why don't well-educated adults understand accumulation? A challenge to researchers, educators, and citizens. *Organizational Behavior and Human Decision Processes, 108*, 116-130. doi:10.1016/j.obhdp.2008.03.003

De Dreu, C. K. W. (2007). Cooperative outcome interdependence, task reflexivity, and team effectiveness: a motivated information processing perspective. *Journal of Applied Psychology, 92*, 628-638. doi:10.1037/0021-9010.92.3.628

DeChurch, L. A., & Haas, C. D. (2008). Examining team planning through an episodic lens: Effects of deliberate, contingency, and reactive planning on team effectiveness. *Small Group Research, 39*, 542-568. doi:10.1177/1046496408320048

DeChurch, L. A., & Mesmer-Magnus, J. R. (2010). The cognitive underpinnings of effective teamwork: a meta-analysis. *Journal of Applied Psychology, 95*, 32-53. doi:10.1037/a0017328

Fisher, D. M. (2014). Distinguishing between taskwork and teamwork planning in teams: Relations with coordination and interpersonal processes. *Journal of Applied*

Psychology, 99, 423-436. doi:10.1037/a0034625

Fulmer, C. A., & Ostroff, C. (2016). Convergence and emergence in organizations: An integrative framework and review. *Journal of Organizational Behavior, 37*, 122-145. doi:10.1002/job.1987

Gabadinho, A., Ritschard, G., Müller, N. S., & Studer, M. (2011). Analyzing and visualizing state sequences in R with TraMineR. *Journal of Statistical Software, 40,* 1-37. doi:10.18637/jss.v040.i04

Gersick, C. J. (1991). Revolutionary change theories: A multilevel exploration of the punctuated equilibrium paradigm. *Academy of Management Review, 16,* 10-36. doi:10.5465/AMR.1991.4278988

Gevers, J. M., van Eerde, W., & Rutte, C. G. (2001). Time pressure, potency, and progress in project groups. *European Journal of Work and Organizational Psychology, 10,* 205-221. doi:10.1080/13594320143000636

Gurtner, A., Tschan, F., Semmer, N. K., & Nägele, C. (2007). Getting groups to develop good strategies: Effects of reflexivity interventions on team process, team performance, and shared mental models. *Organizational Behavior and Human Decision Processes, 102,* 127-142. doi:10.1016/j.obhdp.2006.05.002

Humphrey, S. E., & Aime, F. (2014). Team microdynamics: Toward an organizing approach to teamwork. *The Academy of Management Annals, 8*, 443–503. doi:10.1080/19416520.2014.904140

Lang, J. W., & Bliese, P. D. (2009). General mental ability and two types of adaptation to unforeseen change: applying discontinuous growth models to the task-change paradigm. *Journal of Applied Psychology, 94*, 411-428. doi:10.1037/a0013803

Konradt, U., Schippers, M. C., Garbers, Y., & Steenfatt, C. (2015). Effects of guided reflexivity and team feedback on team performance improvement: The role of team

regulatory processes and cognitive emergent states. *European Journal of Work and Organizational Psychology, 24*, 777–795. doi:10.1080/1359432X.2015.1005608

Kostka, G., & Fiedler, J. (2016). *Large Infrastructure Projects in Germany*. Chan, Switzerland: Springer International Publishing AG. doi:10.1007/978-3-319-29233-5

Kozlowski, S. W. J., & Klein, K. J. (2000). A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel Theory, Research and Methods in Organizations: Foundations, Extensions, and new Directions* (pp. 3-90). San Francisco, CA: Jossey-Bass.

Marks, M. A., Mathieu, J. E., & Zaccaro, S. J. (2001). A temporally based framework and taxonomy of team processes. *Academy of Management Review, 26*, 356–376. doi:10.2307/259182

Mathieu, J., Maynard, M. T., Rapp, T., & Gilson, L. (2008). Team effectiveness 1997-2007: A review of recent advancements and a glimpse into the future. *Journal of Management, 34*, 410-476. doi:10.1177/0149206308316061

Maynard, M. T., Kennedy, D. M., & Sommer, S. A. (2015). Team adaptation: A fifteen-year synthesis (1998–2013) and framework for how this literature needs to "adapt" going forward. *European Journal of Work and Organizational Psychology, 24,* 652-677. doi:10.1080/1359432X.2014.1001376

Schippers, M. C., Homan, A. C., & van Knippenberg, D. (2013). To reflect or not to reflect: Prior team performance as a boundary condition of the effects of reflexivity on learning and final team performance. *Journal of Organizational Behavior, 34,* 6–23. doi:10.1002/job.1784

Schippers, M. C., West, M. A., & Dawson, J. F. (2015). Team reflexivity and innovation: The moderating role of team context. *Journal of Management, 41*, 769–788.

doi:10.1177/0149206312441210

Shiyko, M. P., Ram, N., & Grimm, K. J. (2012). An overview of growth mixture modeling: A simple nonlinear application in OpenMx. In R. H. Hoyle (Ed.), *Handbook of Structural Equation Modeling* (pp. 532-546). New York: Guilford Press.

Smith, K. G., & Locke, E. A. (1990). Goal setting, planning, and organizational performance: An experimental simulation. *Organizational Behavior and Human Decision Processes, 46*, 118-134. doi:10.1016/0749-5978(90)90025-5

Tannenbaum, S. I., & Cerasoli, C. P. (2013). Do team and invidiaul debriefs enhance performance? A meta-analysis. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 55*, 231-245. doi:10.1177/0018720812448394

Vancouver, J. B., & Weinhardt, J. M. (2012). Modeling the mind and the milieu: Computational modeling for micro-level organizational research. *Organizational Research Methods, 15*, 602-623. doi:10.1177/1094428112449655

Wagner, D., & Disparte, D. (2016). Risk Management in a Global World. In D. Wagner & D. Disparte (Eds.), *Global Risk Agility and Decision Making* (pp. 3-16). London: Palgrave Macmillan UK.

Appendix A

Study 1: Correlations among behavioural indicators and the four dimensions of the REMINT at t_1

Indicators	IS Quantity	IS Quality	IE Quantity	IE Quality
Meetings with study group	.30**	–.02	.02	–.05
Weeks that passed since started to study	.16	.16	.35**	.35**
Duration of daily study	.21*	.10	.30**	.27*
Private coaching	.05	.22*	.19*	.22*
Attendance at tutorials	.23*	.10	.23*	.16
Attendance at lectures	.09	.09	–.04	.02
Initiating forum discussions	.35**	.31**	.32**	.41**

Note. $N = 75$. IS = Information seeking. IE = Information evaluation. No significant correlations were found at t_2 and t_3.

*$p < .05$. **$p < .01$

Appendix B

Study 3: Overview of estimates and corresponding values in model 1 and model 2

Model	Relationship	Estimate	Posterior S.D.	One-tailed p-value	Lower 5%	Upper 5%
1	**T1 IS Qn. – T1 Perf.**	**-0.474**	**0.193**	**0.012**	**-0.768**	**-0.133**
	T1 IS Ql. – T1 Perf.	**0.545**	**0.190**	**0.005**	**0.206**	**0.831**
	T1 IS Qn. – T2 Perf.	-0.200	0.228	0.197	-0.568	0.182
	T1 IS Ql. – T2 Perf.	0.152	0.225	0.253	-0.225	0.515
	T2 IS Qn. – T2 Perf.	0.018	0.208	0.467	-0.321	0.364
	T2 IS Ql. – T2 Perf.	0.115	0.204	0.290	-0.229	0.445
	T2 IS Qn. – T3 Perf.	-0.155	0.176	0.197	-0.434	0.146
	T2 IS Ql. – T3 Perf.	-0.227	0.212	0.146	-0.570	0.126
	T3 IS Qn. – T3 Perf.	0.060	0.151	0.346	-0.192	0.306
	T3 IS Ql. – T3 Perf.	**0.433**	**0.172**	**0.009**	**0.136**	**0.700**
2	T1 IE Qn. – T1 Perf.	-0.141	0.207	0.253	-0.474	0.206
	T1 IE Ql. – T1 Perf.	**0.274**	**0.204**	**0.099**	**-0.077**	**0.594**
	T1 IE Qn. – T2 Perf.	-0.062	0.225	0.393	-0.431	0.309
	T1 IE Ql. – T2 Perf.	**-0.289**	**0.218**	**0.100**	**-0.634**	**0.083**
	T2 IE Qn. – T2 Perf.	-0.102	0.225	0.330	-0.462	0.279
	T2 IE Ql. – T2 Perf.	**0.497**	**0.235**	**0.024**	**0.086**	**0.859**
	T2 IE Qn. – T3 Perf.	-0.019	0.210	0.463	-0.360	0.330
	T2 IE Ql. – T3 Perf.	0.053	0.235	0.411	-0.339	0.432
	T3 IE Qn. – T3 Perf.	0.048	0.193	0.403	-0.272	0.364
	T3 IE Ql. – T3 Perf.	-0.029	0.204	0.445	-0.361	0.313

Note. IS = Information seeking. IE = Information evaluation. Perf = Performance. Estimates below the significance level of .1 in bold.

Appendix C

Study 3: Overview of estimates and corresponding values in model 3 and model 4

Model	Relationship	Estimate	Posterior S.D.	One-tailed p-value	Lower 2.5%	Upper 2.5%
3	T4 IS Qn. – T4 Perf.	0.027	0.169	0.437	-0.254	0.303
	T4 IS Ql. – T4 Perf.	0.184	0.168	0.143	-0.101	0.450
	T4 IS Qn. – T5 Perf.	0.022	0.194	0.455	-0.297	0.341
	T4 IS Ql. – T5 Perf.	**-0.294**	**0.169**	**0.045**	**-0.565**	**-0.008**
	T5 IS Qn. – T5 Perf.	0.149	0.196	0.230	-0.183	0.464
	T5 IS Ql. – T5 Perf.	**0.335**	**0.202**	**0.051**	**-0.002**	**0.664**
	T5 IS Qn. – T6 Perf.	**-0.315**	**0.217**	**0.080**	**-0.660**	**0.053**
	T5 IS Ql. – T6 Perf.	0.152	0.226	0.253	-0.225	0.521
	T6 IS Qn. – T6 Perf.	**0.244**	**0.164**	**0.075**	**-0.036**	**0.504**
	T6 IS Ql. – T6 Perf.	-0.025	0.173	0.443	-0.308	0.262
4	T4 IE Qn. – T4 Perf.	-0.066	0.167	0.347	-0.340	0.212
	T4 IE Ql. – T4 Perf.	**0.299**	**0.163**	**0.041**	**0.017**	**0.554**
	T4 IE Qn. – T5 Perf.	**-0.262**	**0.188**	**0.087**	**-0.564**	**0.055**
	T4 IE Ql. – T5 Perf.	**0.271**	**0.161**	**0.051**	**-0.002**	**0.528**
	T5 IE Qn. – T5 Perf.	**0.522**	**0.189**	**0.005**	**0.193**	**0.814**
	T5 IE Ql. – T5 Perf.	**-0.286**	**0.189**	**0.072**	**-0.585**	**0.035**
	T5 IE Qn. – T6 Perf.	0.037	0.190	0.423	-0.277	0.348
	T5 IE Ql. – T6 Perf.	-0.223	0.186	0.121	-0.518	0.093
	T6 IE Qn. – T6 Perf.	-0.048	0.192	0.402	-0.361	0.272
	T6 IE Ql. – T6 Perf.	0.241	0.188	0.106	-0.079	0.540

Note. IS = Information seeking. IE = Information evaluation. Perf = Performance. Estimates below the significance level of .1 in bold.

Lebenslauf Kai-Philip Otte

Nationalität	Deutsch
Geburtsdaten	01.10.1987, Eckernförde
Anschrift	Schauenburgerstraße 110a
	24118 Kiel
Kontakt	otte@psychologie.uni-kiel.de

Ausbildung	06/2017	Erlangung des Doktorgrades der Psychologie an der CAU zu Kiel
	2013-2017	Promotionsstudium im Fach Psychologie an der CAU zu Kiel
	07/2013	Hochschulabschluss: Diplom in Psychologie
	2008-2013	Psychologiestudium CAU zu Kiel
	07/2007	Schulabschluss: Abitur
	1999-2007	Stiftung Louisenlund