

[In *Inquiry: Critical Thinking Across the Disciplines* 30:1, 40 – 50, 2015]

**Measuring Critical Thinking About Deeply Held Beliefs:  
Can the California Critical Thinking Dispositions Inventory Help?**

*Ilan Goldberg, Justine Kingsbury, Tracy Bowell, and Darelle Howard*

**Abstract**

The California Critical Thinking Dispositions Inventory (CCTDI) is a commonly used tool for measuring critical thinking dispositions. However, research on the efficacy of the CCTDI in predicting good thinking about students' own deeply held beliefs is scant. In this paper we report on our study that was designed to gauge the usefulness of the CCTDI in this context, and take some first steps towards designing a better method for measuring strong sense critical thinking.

Key words: critical thinking, critical thinking dispositions, critical thinking tests, CCTDI, deeply held beliefs, strong sense critical thinking.

**I. Introduction**

The main goal of general critical thinking instruction is to improve the quality of students' reasoning on topics not dealt with directly in the critical thinking class, and in situations other than the critical thinking classroom. Getting students to think critically about their *deeply held beliefs* is recognized as particularly difficult because critical thinking skills, selectively applied, enable students to defend these beliefs from criticism and attack incongruent beliefs (Paul, 1992, p. 136; 1993, pp. 57-8, 137-8, 206-7, 329-30).

By "deeply held beliefs" we mean beliefs that have one or more of the following features:

- They are held passionately, strongly, with a high degree of emotional intensity.
- They are central to the believer's worldview.
- They are held dogmatically: that is, they are not readily abandoned or revised in the light of new evidence.

The belief that Jesus is divine, the belief that people who vote for the left don't understand politics, and the belief that eating meat is morally okay would all be examples provided they were sufficiently important to the participant. The participant need not be aware of holding such beliefs, or of holding them strongly, so long as it can be shown that she behaves as if she does.

The ideal that critical thinking instruction aims at is what Richard Paul (Paul 1993, pp. 137-8) calls "strong sense critical thinking" (SSCT): critical thinking that is not biased in favor of one's own beliefs, but is applied as stringently to one's own beliefs as it is to the beliefs of one's opponents. Paul contrasts this with "weak sense critical thinking" (WSCT): "critical thinking skills internalized in the service of one's vested interests and desires" (Paul 2000, p. 258). Moreover, Brookfield (2003), Govier (1999, pp. 248–50), Mezirow (1990, p. 4), and Segal (1998) illustrate the connection between the stresses involved in having one's beliefs criticized and uncritical defensive reactions.

(See also Paul (1993, pp. 57–8, 137, 206–7, 329–30), Adler (1994, p. 108), Hahn (1992, p. 165), Irving & Williams (1995), Johnson (1992, pp. 73–4), and Kornblith (1999, p. 183).

One of the keys to the success of a critical thinking intervention is imparting the *will* to think critically – a general disposition to think critically when critical thinking is called for – in addition to the *ability* to do so. Once a student decides to think critically about a topic, additional, more specific critical thinking dispositions come into play. The disposition to evaluate one’s own and alternative views by the same standards and the disposition to persevere through a prolonged or difficult thinking process, amongst others, determine the quality of students’ thinking once it has been switched on. Because of the importance of people’s deeply held beliefs in determining how they interact with others in the wider world, the disposition to think about such beliefs critically, and attendant dispositions that support good thinking, are particularly important both to promote and, ipso facto, to measure.

## II. The Project

### *The Instruments*

The California Critical Thinking Dispositions Inventory (CCTDI) is a tool that is commonly used to measure critical thinking dispositions in the field. See for example Bartlett & Cox (2002), Giddens & Gloeckner (2005), Hicks-Moore & Pastirik (2006), Nieto & Valenzuela (2012), Rudd & Hoover (2000), Sankaran & Dimitrijevic (2010), Stone et al. (2001), Ip et al. (2000). However, research on the efficacy of the CCTDI in predicting good thinking about students’ own deeply held beliefs is scant. The main aim of the present study was to gauge the usefulness of the CCTDI in measuring the quality of students’ thinking about their own deeply held beliefs. This was done through comparison of CCTDI results with results from two other measures: the Revised Paranormal Belief Scale (RPBS, Tobacyk, 2004) and a score from a semi-structured interview. The outcome is important for a further study we hope to undertake that will compare alternative methods of delivery of the general critical thinking course to see which is most successful in promoting critical thinking about deeply held beliefs, that is, genuine strong sense critical thinking (SSCT). This future study will require us to measure the quality of students’ thinking about these beliefs; hence our interest in determining what is a good way to measure it.

### *The First Hypothesis*

One of our hypotheses was that CCTDI results would *not* correlate so well with the interview score for SSCT, because the CCTDI appears to measure test-takers’ *opinions* about some of their attitudinally relevant beliefs, values and intentions, rather than the dispositions themselves, and there is reason to think that judgments of these sort tend to be more positive than is justified. (See for example Helzer & Dunning (2012), pp. 379-396).

### *The Second Hypothesis*

A second hypothesis was that lack of SSCT is indicated by the presence of any one deeply held belief which is pseudoscientific or anti-scientific, or which lacks good scientific backing even though it is a belief that could reasonably be expected to have

such backing if it were true. The RPBS is a rough and ready survey of seven popular kinds of beliefs of this sort (e.g. witchcraft, precognition). We hypothesised that there would be a correlation between *highest* score on the RPBS and *low* scores for SSCT (as measured by scored interviews). We surmised this because all our participants had completed a general critical thinking course, and our expectation was that participants who were inclined towards thinking critically in the strong sense as shown by their high SSCT scores and who had powerful critical thinking tools at their disposal, would show a lower level of belief in paranormal items as measured by the RPBS. Likewise, participants who were not disposed to think critically in the strong sense as shown by low SSCT scores would score higher on the RPBS, because of their lack of interest in thinking critically about their own controversial beliefs and because of their new ability to fend off the critiques of others.

Had it turned out that the RPBS scores performed better as predictors of individual participants' SSCT scores than the CCTDI, that would have been a useful result for us, since the RPBS is free, quick to administer, and can be easily modified, unlike the CCTDI. In fact, however, neither the CCTDI nor the RPBS performed sufficiently well in the present study as predictors of individual participants' SSCT for us to be confident about using them in our future research. Although some correlations were found, all were judged to be insufficient. We also found a surprisingly large gap between CCTDI scores (where scores were relatively good), and our interview results (where most participants did not score well). The RPBS performed better than the CCTDI in that respect. When taken as a single group, our participants showed a considerable lack of inclination to think critically about their own deeply held beliefs, as well as a high degree of belief in the kinds of paranormal statements included in the RPBS. This suggests that measuring beliefs in some epistemologically problematic claims might prove valuable in the development of an objective measure of SSCT. However, we have decided to use a modified version of our own semi-structured interview method in our own future research.

### III. Method

There were 57 participants, all of whom had completed an introductory-level one-semester course in critical thinking at the [university omitted from blind review copy] during the three years preceding the study. Participants were given two cinema vouchers as compensation for their efforts. Data from all participants was of sufficient quality to be included in the analysis. At this initial stage no demographic data was collected. Data was collected using two paper and pencil questionnaires and a semi-structured interview.

The first questionnaire given was the Revised Paranormal Belief Scale (RPBS). This is a standardized measure of degree of belief in each of seven kinds of "paranormal" beliefs: Traditional Religious Belief, Psi, Witchcraft, Superstition, Spiritualism, Extraordinary Life Forms, and Precognition. It is comprised of 26 statements (e.g. "Astrology is a way to accurately predict the future;" "The soul continues to exist though the body may die;" "Witches do exist;" "A person's thoughts can influence the movement of a physical object"). Participants indicate the strength of their agreement/disagreement with each statement on a seven-level Likert-like rating scale (1 = strongly disagree, 7 = strongly agree). Scores for each of the seven kinds of paranormal belief were calculated by averaging the scores for three to four related statements.

Highest RPBS – the highest degree of belief in any one of the seven categories scored by each participant – was also calculated.

After completing the RPBS, participants were given thirty minutes to complete the California Critical Thinking Dispositions Inventory (CCTDI). This instrument has been used internationally in research, education and business to measure critical thinking dispositions. It is comprised of 75 statements. To complete the inventory, participants need to indicate the strength of their agreement/disagreement on a six-level Likert-like rating scale. Participants' responses were machine scored by Insight Assessment, who own and distribute the CCTDI. Every participant was given a score for each of the following characteristics: Truth-seeking, Open-mindedness, Inquisitiveness, Analyticity, Systematicity, Confidence in Reasoning, and Maturity of Judgment. They were also given a mean score, which is as close as the CCTDI gets to measuring a general or overall disposition towards critical thinking. A CCTDI score of "40 points or higher indicates a positive inclination or affirmation of the characteristic" (Giancarlo & Facione, 2001, p. 37).

Next came a three-part semi-structured interview process to determine a strong sense critical thinking score (SSCT). After taking the CCTDI participants were interviewed by one of the researchers. One researcher was the main interviewer, conducting 55 of the 57, while a second researcher conducted two of the interviews. The interview normally took twenty to thirty minutes and was divided into three parts, each geared towards answering a different question. First, participants were asked to rate the influence of the critical thinking course on the quality of their thinking in general (Interview Question 1, Q1 hereafter). Second, participants were asked to rate the influence of the critical thinking course on the quality of their thinking about their own deeply held beliefs (Interview Question 2, Q2 hereafter). The meaning of "deeply held beliefs" was explained to participants briefly in a way that is consistent with but not necessarily as detailed as that given above. Both these questions were later scored by each of the researchers scoring independently on a seven-level Likert-like rating scale (1 = a very negative effect, 7 = a very positive effect). If a participant's answer was not sufficiently clear, he or she was explicitly asked to rate changes on this scale.

In the third part of the interview (Interview Question 3, Q3 hereafter), the participant and the interviewer engaged in a discussion about one or more of the participant's controversial deeply held beliefs ("controversial" was explained as "such that intelligent and educated people might disagree about it"). Generally two deeply held beliefs were discussed, but in some interviews discussing one belief provided enough information about the participant's ability and inclination towards SSCT, and in other interviews three or more were needed. Typically the first belief to be discussed was chosen from amongst statements the participant rated six or seven in the RPBS. Having discussed one such claim with the participant in detail, the interviewer would typically ask the participant to volunteer a deeply held controversial belief that would be suitable as a second topic for discussion. In many cases the participants did not come up with a suitable belief for discussion: consequently many of the interviews used one chosen by the interviewer, "It is morally okay to eat meat."

For each belief discussed, participants were first asked to give the reasoning behind their belief. They were then presented with arguments that undermined their reasons, and asked to volunteer their thoughts again. The process was normally repeated

twice for each belief discussed. To help elicit the needed data participants were also asked questions like these:

- What are your main reasons for thinking that P is true/false?
- Suppose you met someone who had different beliefs about P, because they think that Q or R. How do you think you might respond to them about this issue?
- Has this discussion changed your view about P?
- In the light of this discussion, do you think that in the days and weeks following the interview you will think about P further, and perhaps try to find out more information about the topic?

Then, four raters, two of whom were the original interviewers, used audio recordings of the interviews and a scoring guide, to independently rate the behaviour of each participant on a seven-level Likert-like rating scale for its conformity with the following claim: “The participant displays strong sense critical thinking about his/her own deeply held controversial beliefs.”

The scoring guide included definitions of SSCT versus WSCT or non-critical thinking, as well as lists of typical behaviours indicative of these (see Appendix). The guide also defined a scale progressing from 1 (All the behaviours exhibited conform or are sufficiently close to the definition and examples of WSCT or non-critical thinking) to 7 (All the behaviours exhibited conform or are sufficiently close to the SSCT definition and list of examples).

Interview Question 3 was the interview question we were most interested in. We hoped that there might be sufficiently high correlations between a participant’s SSCT, as judged by the researchers on the basis of Q3 responses, and results on either the RPBS or the CCTDI, such that in future we could use one of these standardised tests rather than the time-consuming interview method to determine how critically students are thinking about their own deeply held beliefs, as well as the level of their inclination to do so.

#### **IV. Results**

On the RPBS, scores for each of seven kinds of paranormal belief were calculated for each participant by averaging their scores for three or four related statements. We were particularly interested in each participant’s highest RPBS (the highest of these seven scores). For nine participants, the highest RPBS was 7; for twelve participants the highest RPBS was between 6 and 7, tailing off gradually, while at the other end of the scale three participants had a highest RPBS of 1. The mean for the highest RPBS was 4.65, indicating substantial paranormal belief in general among the 57 participants.

The mean CCTDI score of 49 of the 57 participants indicated “a positive inclination or affirmation” (Giancarlo & Facione, 2001) towards critical thinking. The scores of the other eight indicate “ambiguity or ambivalence toward the characteristic” (Ibid, 2001); none of the participants were negatively disposed toward critical thinking according to this measure. The CCTDI means of all the participants for each of the seven categories are either in the positive range (above a value of 40), or in the case of truth-seeking (mean of 38) and systematicity (mean of 39.8), very close to it. That is to say, our participants scored rather well on the CCTDI. The average for Mean CCTDI was 43.77.

In response to Interview Question 1, 53 of the 57 participants thought that the critical thinking course had a positive effect on the quality of their thinking. The

responses to Interview Question 2 were quite different: only 32 of the participants thought that the course had no effect on the quality of their thinking about their deeply held beliefs, whereas 25 thought it had a positive effect on their thinking about such beliefs. The means for Q1 and Q2 are 5.80 and 4.84 respectively.

Interview Question 3 was intended as a direct measure of participants' SSCT. We averaged the scores given to a participant by the four raters. The highest score, reflecting the participant who displayed the greatest amount of SSCT, was 5.75 (out of a possible 7). The average across all participants was 2.25, indicating widespread deficiency in SSCT. 32 of the 57 participants scored between 1 and 2 inclusive, showing that over half of the participants exhibited a large majority of behaviours that were outlined in our scoring guide as being examples of WSCT. Most significant of all, only six participants scored 4 or above (a 4 was given when the participant's behaviours were evenly divided between strong sense and weak sense critical thinking behaviours). Of those six, only two scored above 5.

Table 1: RPBS, CCTDI and Interview Means

	Mean	Standard Deviation
RPBS Traditional Religious Belief	3.57	2.16
RPBS Psi	2.70	1.53
RPBS Witchcraft	3.12	1.88
RPBS Superstition	1.59	1.04
RPBS Spiritualism	3.32	1.83
RBPS Extraordinary Life Forms	2.91	1.26
RBPS Precognition	2.82	1.35
RPBS Highest	4.65	1.937
CCTDI Truth Seeking	38.00	5.57
CCTDI Open-mindedness	43.82	5.74
CCTDI Inquisitiveness	49.05	5.49
CCTDI Analyticity	45.49	5.32
CCTDI Systematicity	39.84	5.87
CCTDI Confidence in Reasoning	44.75	6.06
CCTDI Maturity of Judgment	45.70	6.05
CCTDI Mean	43.77	3.49
Interview Question 1	5.80	0.79
Interview Question 2	4.84	0.86
Interview Question 3	2.25	1.13

The scoring of Q3 required raters to identify the behaviours of the participants as behaviours characteristic of SSCT or behaviours characteristic of WSCT. There were significant positive correlations between the four raters: correlations between pairs of raters range from .572 to .762. We had hoped for inter-rater reliability of .8 or above: however, the inter-rater correlations are high enough to show the interview-rating method has potential as a measure of SSCT.

Table 2: Inter-rater Pearson Correlations, Interview Q3

	Rater 1	Rater 2	Rater 3
Rater 2	.762**		
Rater 3	.670**	.608**	
Rater 4	.595**	.594**	.572**

\*\* . Correlation is significant at the 0.01 level (2-tailed).

\* . Correlation is significant at the 0.05 level (2-tailed).

There was a significant negative correlation between the highest RPBS scores and SSCT scores for three out of four raters, although it did not correlate significantly for the mean of all four raters. However, the negative correlation of SSCT scores with the mean for the highest RPBS scores was very close to being a significant correlation. There was a significant positive correlation between Mean CCTDI and SSCT as scored by two of the raters, and between Mean CCTDI and Mean SSCT. The highest correlation between Mean SSCT and any other individual item was with CCTDI Inquisitiveness (.392). Curiously, this was also the CCTDI item where our participants scored highest (Mean 49.05).

Table 3: RPBS and CCTDI Pearson Correlations with Interview Q3

	Rater 1	Rater 2	Rater 3	Rater 4	Mean
RPBS Traditional Religious Belief	-.289*	-.255	-.339**	-.033	-.232
RPBS Psi	-.173	-.163	-.244	.080	-.126
RPBS Witchcraft	-.122	-.158	-.241	.101	-.091
RPBS Superstition	-.156	-.204	-.145	.070	-.113
RPBS Spiritualism	-.148	-.112	-.115	.206	-.030
RPBS Extraordinary Life Forms	-.045	-.095	.050	.286*	.058
RPBS Precognition	-.207	-.229	-.151	.111	-.133
RPBS Highest	-.316*	-.298*	-.297*	.035	-.228
CCTDI Truth-seeking	.276*	.360**	.091	.113	.236
CCTDI Open-mindedness	.252	.288*	.294*	.229	.308*
CCTDI Inquisitiveness	.355**	.501**	.252	.254	.392**
CCTDI Analyticity	.246	.284*	.120	-.026	.161
CCTDI Systematicity	.112	.058	-.088	-.061	.001
CCTDI Confidence in Reasoning	.163	.265*	-.047	.106	.152
CCTDI Maturity of Judgment	.089	.188	.056	-.077	.071
CCTDI Mean	.343**	.446**	.147	.118	.301*

\*\* . Correlation is significant at the 0.01 level (2-tailed).

\* . Correlation is significant at the 0.05 level (2-tailed).

## V. Discussion

Although we hypothesised a mismatch between Q3 and CCTDI results, due to the influence of self-reporting in the latter, the extent of the difference seems to us to be

problematic. The mostly good CCTDI results indicated that our participants were by and large disposed towards thinking critically. Q3 results showed that this was not the case, at least where participants' own deeply held beliefs were involved. This provides additional empirical support for the existence of the problem of weak sense critical thinking. The same can be said about the differences between Q1 and Q2 results: While almost all participants claimed that the critical thinking course improved the quality of their thinking, most thought that the course had no effect on the quality of their thinking about their deeply held beliefs. Of those who did claim some improvement, several thought that the course helped them clarify and strengthen their prior beliefs, and that it helped them improve the quality of their responses to others in arguments.

The problem of WSCT, then, may be an important underlying similarity explaining the disparities between Q1 and Q2 results, and those between the CCTDI and Q3 results. Given the importance of critical thinking dispositions in teaching for strong sense critical thinking, as attested by Paul (1993, pp. 325-327) and Siegel (1988, p. 39), and given that a majority of participants readily admitted lack of transfer to their own deeply held beliefs, one may rightly ask whether the CCTDI is scored rather optimistically. This was particularly evident with CCTDI Inquisitiveness, where our participants' mean was remarkably close to the high range (49.05). According to the CCTDI Test Manual (Facione & Facione, 2010, p. 20), scores in the 40-50 range "indicate consistent endorsement and valuation of the attitude," and scores in the 50-60 range "indicate that the attribute or attitude is a positive habit of mind and likely to factor into the individual's approach to all higher order thinking (reflective problem definition and problem solving), particularly when the situation is of high consequence." Levels of the self-directed inquisitiveness that is so important for SSCT – the kind that prompts us to question our own fundamental assumptions – do not seem to be in line with the CCTDI scale.

It was hoped that the scoring of Q3 would provide an accurate measure of participants' SSCT, acting as a standard with which to evaluate the suitability of the CCTDI and the RPBS for future research. However, in order to be completely confident in our use Q3 for this purpose, the correlations between the scores given by the different raters would need to be .8 or above, and even the best-correlated pair of raters do not quite meet this standard. This means that the significant but low correlations found between Q3 and the RPBS may well increase, once our interview and scoring methods improve. The approach taken by the RPBS may also yield better results if the kinds of beliefs measured were specifically tailored to include all the major paranormal beliefs currently popular with students in [country omitted from blind review copy] (our broad demographic). Therefore, measuring degree of belief in some epistemologically problematic claims could still have a role to play in the development of a standardised measure of SSCT. A somewhat similar approach is taken in Williams, Aspiranti & Krohn (2010) using political beliefs.

The lack of sufficiently high inter-rater reliability in the scoring of SSCT does not have a similar import with regards to the suitability of the CCTDI. The largely positive CCTDI results and the mostly negative Q3 results mean that the CCTDI cannot serve as a proxy for a more direct measure of SSCT even if the low correlations between the two were somewhat increased.



With a view to improving inter-rater reliability for the next stage of our research, we conducted a qualitative analysis of discrepant Q3 ratings in order to identify possible sources of these discrepancies between the raters. Five were identified:

1. Answers to Q1 (What effect does the participant think the course had on the quality of his/her thinking?) and Q2 (What effect does the participant think the course has on the quality of his/her thinking about his/her own deeply held beliefs?) were known to the raters at the time of rating. This probably influenced the rating of Q3 in some cases, especially considering the overlap between Q2 and Q3. In future iterations Q1 and Q2 will be included in a questionnaire that will not be accessible to the raters.
2. Another possible source of discrepancies in the scoring of Q3 arises from differing views about whether the quality of the initial arguments given by the participant in support of their view was to be taken into consideration. After the belief to be discussed was decided, the interviewer posed a question such as “Can you give me your main reasons for believing that P?”; to which the interviewee responded. Sometimes these initial arguments were very poor. Rater 3 took the quality of these initial arguments into consideration when rating Q3. The participants were asked to provide good reasons, or their best reasons, for the beliefs under discussion. On Rater 3’s view, if they put forward a bad argument without noticing the problems with it or revising it, then this displayed a lack of SSCT. Raters 2 and 4 did not take the quality of the initial arguments into consideration. Some of the topics discussed were not ones the participants had given much prior thought to, and, Raters 2 and 4 thought, it was their responses to the interviewer’s counterarguments that revealed the quality of the interviewee’s thinking. For the most part Rater 1 focussed on the counterarguments as the basis for assessment: however if the initial arguments presented were especially poor, Rater 1 did note this down as a possible signal for behaviours to come. This discrepancy could be solved by either ignoring initial arguments uniformly, or by asking the participants more forcefully to give only arguments that they consider good, while giving them a few moments to gather their thoughts.
3. The raters were not always neutral with respect to the controversial beliefs discussed. We cannot completely discount the possibility that our own opinions influenced our rating, although we certainly did our best to avoid such influences.
4. Two members of the team took on a dual role of both interviewers and raters. One way in which this might have had an influence on the rating of Q3 is that interviewers had more information about the participants, being in a position to notice their body language, for example; another is that the interviewers might have over-rated the quality of their own arguments, and underrated the participant’s responses. These roles will need to be separated.
5. Lastly – and this may have been the root cause of all of the above – only three interviews were rated together at the start of the rating period. We believe that comparing ten interviews and ratings at the outset would have been much more likely to help us identify these potential problems and calibrate our rating accordingly. (Thanks to Michael Scriven for pointing this out.)

## **VI. Conclusion**

Neither the CCTDI nor the RPBS were shown to be useful as a proxy for a more direct measure of SSCT about deeply held beliefs, although the general approach taken by the RPBS may still be of some use. We see no overwhelming obstacle to reaching the desired level of inter-rater reliability, and so our semi-structured interview, and Q3 in particular, seems most promising for our project as a potential direct measure of how critically people are thinking about their deeply held beliefs. We hope in future research to modify our interviewing and interview-scoring methods in order to achieve a level of inter-rater reliability that will make some variant of Q3 a credible direct measure of SSCT. Teaching for strong sense critical thinking about deeply held beliefs is an important goal for critical thinking educators, and developing pedagogical methods that achieve this goal requires being able to accurately measure how critically students are thinking about their deeply held beliefs.

### Acknowledgements

Thanks to Nicola Starkey for her help with interpreting the data and to the Faculty of Arts and Social Sciences at the University of Waikato for funding this research. Thanks also to the University of Waikato Summer Research Scholarship Programme for providing Darelle Howard the opportunity to participate in writing this paper. Finally thanks to the audience at OSSA 10 and in particular Robert Ennis for valuable feedback.

### References

- Adler, Jonathan E. (1994). Self-criticism: praise, problems, and problems of praise. In Ralph H. Johnson & Anthony J. Blair (Eds.), *New Essays in Informal Logic* (pp. 104–115). Windsor, Ontario, Canada: Informal Logic.
- Bartlett, D. J., & Cox, P. D. (2002). Measuring change in students' critical thinking ability: implications for health care education. *Journal of Allied Health, 31*(2), pp. 64-69.
- Brookfield, Stephen (2003). Critical thinking in adulthood. In *Critical Thinking and Reasoning: Current Research, Theory, and Practice*. (pp. 143–163). Cresskill, NJ: Hampton Press.
- Facione, N. C. & Facione, P. A. (2010). *The California Critical Thinking Dispositions Inventory: CCTDI Instrument User's Manual*. Millbrae, CA: The California Academic Press.
- Facione, P. A., Facione, N. C., & Giancarlo, C. A. (2000). The disposition toward critical thinking: Its character, measurement, and relationship to critical thinking. *Informal Logic, 20*(1), 61-84.
- Giancarlo, C. A., & Facione, P. A. (2001). A look across four years at the disposition toward critical thinking among undergraduate students. *The Journal of General Education, 50*(1), 29-55.
- Giddens, J., & Gloeckner, G. W. (2005). The relationship of critical thinking to performance on the NCLEX-RN. *The Journal of nursing education, 44*(2).
- Govier, Trudy (1999). Progress and regress on the dialectical tier. In John Hoaglund (Ed.), *The Philosophy of Argument* (pp. 223–240). Newport News, Virginia: Vale Press.

- Hahn, Carl R. (1992). Should fallacies be used to teach critical thinking? In Wendy Oxman, Mark Weinstein, & Nicholas M. Michelli (Eds.), *Critical Thinking: Implications for Teaching and Teachers* (pp. 164–169). Upper Montclair, NJ: Institute for Critical Thinking.
- Helzer, E. G., & Dunning, D. (2012). On motivated reasoning and self-belief. In S. Vazire & T. D. Wilson (Eds.), *Handbook of self-knowledge* (pp. 379-396). New York, US: Guilford Press
- Hicks-Moore, S. L., & Pastirik, P. J. (2006). Evaluating critical thinking in clinical concept maps: a pilot study. *International Journal of Nursing Education Scholarship*, 3(1).
- Ip, W. Y., Lee, D. T., Lee, I. F., Chau, J. P., Wootton, Y. S., & Chang, A. M. (2000). Disposition towards critical thinking: a study of Chinese undergraduate nursing students. *Journal of Advanced Nursing*, 32(1), 84-90.
- Irving, J. A. & Williams, D. I. (1995). Critical thinking and reflective practice in counselling. *British Journal of Guidance & Counselling*, 23(1), 107–114.
- Johnson, Ralph H. (1992). Critical reasoning and informal logic. In Richard A. Talaska (Ed.), *Critical Reasoning in Contemporary Culture* (pp. 69–88). Albany: State University of New York Press.
- Kornblith, Hilary (1999). Distrusting reason. *Midwest Studies in Philosophy*, 23, 181–196.
- Mezirow, Jack (Ed). (1990). *Fostering Critical Reflection in Adulthood: A Guide to Transformative and Emancipatory Learning*. San Francisco: Jossey-Bass.
- Nieto, A. M. & Valenzuela, J. (2012). A study of the internal structure of critical thinking dispositions. *Inquiry: Critical Thinking Across the Disciplines*, 27(1), 31-38.
- Paul, R.W. (1992). Teaching critical reasoning in the strong sense: Getting behind worldviews. In R.A. Talaska (Ed.), *Critical Reasoning in Contemporary Culture* (pp. 135-156). Albany: State University of New York Press.
- Paul, R.W. (1993). *Critical Thinking: What Every Person Needs to Survive in a Rapidly Changing World*. (3rd ed.) Santa Rosa, CA: Foundation for Critical Thinking.
- Paul, R.W. (2000). Critical thinking, moral integrity, and citizenship: Teaching for the intellectual virtues. In G. Axtell (ed.), *Knowledge, Belief and Character: Readings in Virtue Epistemology* (pp.255-267). Lanham MD: Rowman and Littlefield.
- Rudd, R., Baker, M., & Hoover, T. (2000). Undergraduate agriculture student learning styles and critical thinking abilities: Is there a relationship? *Journal of agricultural education*, 41(3), 2-12.
- Sankaran, H., & Dimitrijevic, M. (2010). Implications for critical thinking dispositions: Evidence from freshmen in New Mexico. *Inquiry: Critical Thinking Across the Disciplines*, 25(2), 27-35.
- Segal, Steven (1998). Philosophy as a Therapeutic Activity. *Inquiry: Critical Thinking Across the Disciplines*, 17(3), 36–47.
- Siegel, Harvey (1988). *Educating Reason: Rationality, Critical Thinking, and Education*. London: Routledge.
- Stone, C. A., Davidson, L. J., Evans, J. L., & Hansen, M. A. (2001); Validity evidence for using a general critical thinking test to measure nursing students' critical thinking. *Holistic Nursing Practice*, 15(4), 65-74;
- Tobacyk, J.J. (2004). A revised paranormal belief scale. *The International Journal of Transpersonal Studies*, 23, 94-98.

Williams, R. L., Aspiranti, K. B. & Krohn, K. R. (2010). Critical thinking and sociopolitical values reflective of political ideology. *Inquiry: Critical Thinking Across the Disciplines*, 25(3), 22-30.

### **Author Information**

Dr. Ilan Goldberg is a Research Associate in the Philosophy Programme at the University of Waikato; he may be contacted at goldberg@waikato.ac.nz. Dr. Justine Kingsbury is a Senior Lecturer in Philosophy at the University of Waikato, and she may be contacted at justinek@waikato.ac.nz. Dr. Tracy Bowell is also a Senior Lecturer in Philosophy at the University of Waikato, and she can be contacted at taboo@waikato.ac.nz. Darelle Howard is a graduate student in Sociology at the University of Waikato.

### **Appendix: Excerpts From the Interview Scoring Guide**

Strong sense critical thinking is displayed in behaviors like these:

1. Seriously considering opposing arguments when confronted with them.
2. Recognising and internalizing the import of opposing arguments when confronted with them.
3. Showing genuine interest in opposing arguments or evidence.
4. Questioning the perspectives, beliefs, and inferences one is personally invested in.
5. Sympathetically and accurately paraphrasing opposing arguments in a genuine attempt to recognize their strengths and significance.
6. Reasoning from within opposing perspectives in a genuine attempt to recognize their strengths and significance, as well as the limitations of one's own perspective.
7. Showing a genuine interest in basing one's beliefs on the best available evidence.
8. Showing a genuine interest in seeking opposing evidence.

Weak sense critical thinking and non-critical thinking are the opposite of strong sense critical thinking and are displayed in behaviors such as these:

1. Paying lip-service to the import of opposing arguments when confronted with them as a rhetorical device, or a prelude to counter-attack.
2. Lack of interest in opposing arguments or evidence. Lack of inclination to investigate opposing arguments or evidence further. Passing the burden of proof to one's opponents. Thinking that it is reasonable to keep one's deeply held controversial belief until one's opponents supply incontrovertible evidence or proof.
3. Questioning only perspectives, beliefs, and inferences one is not personally invested in, or beliefs and inferences that are contrary or contradictory to those one is personally invested in.
4. Application of the skills taught in the critical thinking course to arguments opposing one's views more than to one's own reasoning.
5. Refusal to reason from within opposing perspectives.
6. Criticising opposing evidence unfairly.

7. Judging the quality of the evidence based on its agreement with one's beliefs.
8. Repeatedly and only defending one's own beliefs and attacking the opposition.
9. An admission that no arguments will ever be good enough to elicit change of belief.
10. Shielding one's beliefs from criticism by claiming that everyone has a right to their own beliefs.

Although we don't test for argument analysis and fallacy identification skills in this research, poor – and only poor – argument analysis and fallacy identification skills are relevant to scoring the third scale of the interview. The reason for this is our use of Richard Paul's three-way distinction between SSCT, WSCT, and unskilled or non-critical thinking. Low dexterity with or use of argument analysis and fallacy identification skills counts against SSCT because SSCT is skilled thinking. But high dexterity with or correct and common use of argument analysis and fallacy identification skills does not count in favour of SSCT because WSCT can be just as skilled as SSCT. It's just that in WSCT, skills are applied apologetically rather than fairmindedly.