

Measuring self-concept of one's own ability with experiment-based behaviour assessment: towards the construct validity of three scoring variants

Martina Frebort¹ & Michaela M. Wagner-Menghin²

Abstract

Self-concept of one's ability is an important construct influencing academic performance (Elliot & Dweck, 2005). The present study deals with the self-concept of one's domain-specific prior knowledge in Psychology. Here, it is measured indirectly through WITE-Psychology (Frebort, unpublished), a knowledge test that measures Psychology student applicants' Psychology knowledge. To measure self-concept of one's domain-specific prior knowledge in Psychology, Wagner-Menghin's (2004) scoring variant was used. However, this scoring variant entailed certain problems such as disadvantaging able testees under some testing conditions and reducing to the overestimation aspect while neglecting underestimation. The present study suggests two new scoring variants ("alternative tendency" score; "sense of reality" score) that overcome the one or the other problematic aspect, respectively. For the original scoring variant ("tendency" score), as well as for the two new scoring variants, unidimensionality is proven through the conditional Likelihood-Ratio test (Andersen, 1973) and graphical model checks. Furthermore, with the correlation coefficient indicating the relationship of each of the three scores to Psychology knowledge, another aspect of construct validity is tested.

Results indicate that for one of the two new scoring variants ("sense of reality" score), items do not cover the latent continuum well; the other new scoring variant ("alternative tendency" score) is one-dimensional and shows the expected independency of Psychology knowledge. For this reason, it is suggested as a promising alternative to the original scoring variant under certain testing conditions.

Key words: verbal academic self-concept; fair scoring; Rasch model; construct validity of LEWITE

¹ Correspondence concerning this article may be addressed to: Martina Frebort, PhD, University of Vienna, Liebiggasse 5, 1010 Vienna, Austria; email: martina.frebort@univie.ac.at

² Department for Medical Education, Medical University of Vienna

Introduction

Because self-concept of one's ability is an important construct that influences achievement performance, its differential diagnostic value in educational and vocational psychology has been observed carefully (Ackerman & Wolman, 2007; Dweck & Leggett, 1988; Elliot & Dweck, 2005; Hong, Chiu, Dweck, Lin & Wan, 1999). An individual who underestimates his or her own abilities may have a lower likelihood of attempting to achieve a valuable goal and thus might miss important opportunities (Ackerman & Wolman, 2007). Kanfer and Ackerman (2005) refer to the self-concept of one's ability as academic self-concept. They conceptualize it as domain-specific and differentiate e.g. math, spatial or verbal domains.

Much research has dealt with the relationship between academic self-concept of one's ability (or self-estimated abilities) and tested abilities. The conclusion is that there is a small (Furnham, 2001; Rammstedt & Rammesayer, 2002) to medium positive relationship between the two (Furnham & Dissou, 2007).

Several available instruments focussing on academic self-concept like the Differential Self-Concept Grid, with a scale measuring self-concept of academic achievement and capabilities (DISK-Gitter mit SKSLF-8, Rost, Sparfeldt, & Schilling, 2007), and the Inventory of Self-Estimated Intelligence (ISI, Rammstedt & Rammesayer, 2002) utilize the questionnaire technique to assess self-concept. However, there is research suggesting that only parts of certain psychologically interesting constructs can be assessed with the questionnaire technique (Kubinger & Litzenberger, 2003). Furthermore, scores derived from personality questionnaires rely on the testee's own estimates; thus they can be distorted in the direction the testees intend (e.g. Kubinger, 2009). For this reason, there has been renewed interest in measures obtained with techniques other than questionnaires.

Creating an experiment-based behaviour assessment (EBA) is one way of assessing a construct without using the questionnaire technique. The EBA-technique takes up the idea of objective personality tests sensu Cattell (1958; Schmidt, 1975), which directly register testees' behaviour in a standardized situation without solely relying on their self-ratings. Scores derived from an EBA take the way an individual behaves while working on an achievement task into account (Kubinger, 2009). EBAs have proven to be incrementally valid beyond personality questionnaires (Kubinger & Litzenberger, 2003).

An EBA-based measure to assess self-concept of one's ability has been suggested by Wagner-Menghin (2004), who modified a multiple choice vocabulary test to observe subjects' self-confidence in their ability while trying to solve the vocabulary achievement task. In the test, subjects are asked to declare for each presented term whether or not they know the term and can explain its meaning. Subsequently, subjects are required to complete a gap-text with two gaps to phrase the term's definition. Four answer options are offered for each gap (see Figure 1). The sum of items a person has declared knowing but has failed to complete in the gap-text forms a measure for self-concept of one's ability ("tendency" score). The sum of correctly defined terms forms a measure for general knowledge ("ability" score). Both scores were scaled using the Rasch model. In addition

to several conventionally administered sets (Frebort, unpublished; see also Frebort & Kubinger, 2008; Litzenberger & Haiden, 2006) an adaptive version of this test (Wagner-Menghin, 2004) exists.

<u>Item type A: Declaring whether one knows a term</u>	
Example: Do you know the following term? alphabet Please mark, wheter you know the significance of the term or not by choosing YES or NO. ☐ YES ☐ NO	
Beispiel: Kennen Sie den folgenden Begriff? Alphabet Bitte geben Sie an, ob Sie die Bedeutung des Begriffes kennen, indem Sie JA bzw. NEIN ankreuzen. ☐ JA ☐ NEIN	
<u>Item type B: Completing a gap-text</u>	
Task 1 out of 25 The halo effect means the tendency _____ to be guided _____ of a person. Possible choices: ☐ of a judge ☐ by the expressed opinion ☐ of a superior ☐ by an offending declaration ☐ of a client ☐ by the overall impression or a salient attribute ☐ of a child ☐ by life-events or inconvenient strokes of fate	
Aufgabe 1 von 25 Unter dem Halo-Effekt versteht man die Tendenz _____, sich _____ einer Person leiten zu lassen. Mögliche Antworten: ☐ eines Richters ☐ von der Meinung ☐ eines Vorgesetzten ☐ von der Beleidigung ☐ eines Klienten ☐ von dem allgemeinen Eindruck oder einem auffälligen Merkmal ☐ eines Kindes ☐ von Lebensereignissen oder unangenehmen Schicksalsschlägen	

Figure 1:
Both item types of WITE-Psychology

However, the scoring of the “tendency” score is problematic in that it solely reflects the aspect of overestimation. To form the “tendency” score, one point is assigned to all items that were declared as known but not solved and zero points for all other conditions (i.e. claiming to know and solving; claiming not to know and solving; claiming not to know and not solving; see Table 1). Although zero indicates that knowledge was not overestimated, one must be aware that this scoring leads to a loss of information. Thus, interpreting a low “tendency” score as a lack of overestimation might not be correct for very able persons, who solve a high percentage of presented gap-text items and consequently do not have the opportunity to show overestimation.

Originally, this suboptimal scoring was implemented in order to allow the estimation of Rasch model parameters with the then available software package, which could handle only a limited number of item/person groups. Since newer and more flexible and powerful Rasch modelling software is now available, the question of how to quantify self-concept of one’s ability using Wagner-Menghin’s (2007) procedure validly is reconsidered.

Aims of the current study

The current study suggests two new scoring variants, one of them addressing the problem of assessing overestimation in groups of higher ability testees as well as in groups of lower ability testees, the other addressing the problem of neglecting the aspect of underestimation in the academic verbal self-concept.

The “alternative tendency” score is calculated, as before, by assigning one point to items that are claimed to be known but actually not solved and zero points to items that are claimed not to be known and actually not solved. A new coding procedure is used for the other two possible constellations (claiming to know and solving; claiming not to know and solving): these solved items are coded as missing values and no longer considered in Rasch model analyses (see Table 1). Since solved items are no longer coded with zero but as missing values the “alternative tendency” score ensures that only items for which overestimation can occur are included in the tendency score. This is a new approach compared to the scoring variant of Wagner-Menghin (2004, 2007).

The “sense of reality” score assigns one point to each item that is declared correctly (no matter whether it is solved or not) and zero points to each item that is declared incorrectly. With this coding, high scorers can be interpreted as persons with a high “sense of reality” (see Table 1). By giving credit for every correct estimate, the “sense of reality” score integrates not only the tendency to overestimate one’s own achievement (like the “tendency” score measures) but also the tendency to underestimate it. This is also an innovation beyond the scoring variant reported by Wagner-Menghin (2004, 2007).

In the present study, the content validity of the two proposed scoring variants is scrutinized by testing whether the items still fit the Rasch model when scored using the new scoring, and by exploring the association between the scores of the scoring variants and the “ability score”, which reflects Psychology knowledge.

Table 1:
Three scoring variants for the self-concept of one's own ability

"Tendency" Score		solved	not solved
	marked as known	0	1
	marked as not known	0	0
"Alternative Tendency" Score		solved	not solved
	marked as known	missing values	1
	marked as not known	missing values	0
"Sense of Reality" Score		solved	not solved
	marked as known	1	0
	marked as not known	0	1

Methods

Hypotheses

Two hypotheses are put forward to address construct validity of the original and the proposed new scoring variants.

We consider the items as fitting the Rasch model, especially when scoring them according to the scoring variant "tendency score", since Wagner-Menghin (2004, 2007) has already been able to show this for some other samples.

Hypothesis 1: The items fit the Rasch model according to three different scoring variants ("tendency" score; "alternative tendency" score; "sense of reality" score).

The dichotomous logistic Rasch model (Rasch, 1960/1980) assumes that the probability for a testee v with an ability ξ_v to solve item i with a difficulty of σ_i is

$$P(+|\xi_v, \sigma_i) = \frac{e^{\xi_v - \sigma_i}}{1 + e^{\xi_v - \sigma_i}}. \text{ The probability of solving an item solely depends on testee ability } \xi_v \text{ and the item difficulty } \sigma_i.$$

To check differential item functioning, median splits were used to form the score groups. Andersen's conditional Likelihood-Ratio test (LRT; Andersen, 1973) was applied. Graphical model checks were plotted to illustrate the distribution of the item parameters and, in case of a significant LRT, to suggest which items needed further scrutinizing. Analyses were performed using the software *eRm* (Mair & Hatzinger, 2006; cf. Poinstingl, Mair, & Hatzinger, 2007), which utilizes the conditional maximum likelihood method to estimate model parameters.

Given that persons' ability score is also relevant in practice and should be interpreted validly, the Rasch model fit of the self-concept scoring variants was tested only for items whose gap-text counterparts also fits the Rasch model. This condition was fulfilled for 20 items (see Table 2).

Further evidence for construct validity was obtained by calculating the correlation coefficients between each of the three scores and the “ability” score. According to the literature on the relationship between ability and academic self-concept, this correlation should be only of small to medium magnitude and positive (Furnham, 2001; Furnham & Dissou, 2007; Rammstedt & Rammsayer, 2002).

Hypothesis 2: The correlation between the “tendency” score and the “ability” score might be of medium magnitude and negative, since able persons have fewer opportunities to score high in the “tendency” scoring variant. There are no previous results with regard to the two new scoring variants, but we hypothesize that the “alternative tendency” score should correlate less strongly with the “ability” score than the “tendency” score does. As concerns the “sense of reality” score, we do not have any assumptions about its relationship to the “ability” score.

Sample

Data stem from the *Viennese Self-Assessment Psychology*® which is offered to applicants of undergraduate studies toward a Bachelor’s or a Master’s degree in Psychology. Self-assessment indicates users work by their own responsibility on a psychological test-battery which was developed deliberately according to the requirements that are made on Psychology students in Vienna (cf. Frebort & Kubinger, 2006). There is neither a duty for the applicants to administer this test-battery nor any consequence for admission to the university. But of course, there is extensive feedback, in particular concerning strengths and weaknesses in the study aimed-for (see for general intentions and problems of self-assessments within university admission e.g. Kubinger, Moosbrugger, Frebort, Jonkisz, & Reiß, 2007). The *Viennese Self-Assessment Psychology*® is an offer provided by the *Centre of Testing and Consulting* of the *Division of Psychological Assessment and Applied Psychometrics* at the Faculty of Psychology. Participants were recruited to participate through the internet, flyers, and a public announcement at the orientation lecture for students interested in studying psychology (Faculty of Psychology, University of Vienna). From the summer of 2005 to the summer of 2007, 407 participants (316 females/91 males) took part in the study; 244 identified themselves as Austrian, 139 as German; 17 said they came from other EU-countries, 7 from non-EU countries. The age of the participants varied from 17 to 56 years, with a median of 20 and 75% being 23 years old or younger.

Instrument

A modification of the General Knowledge test LEWITE (Wagner-Menghin, 2004) was applied: WITE-Psychology (Frebort, unpublished; Frebort & Kubinger, 2008; Sonnleitner, Kubinger, & Frebort, 2009). Each testee worked on 25 items of varied difficulty, administered as a conventional test with fixed item order on the computer.

Procedure

The WITE-Psychology test was completed as the third out of seven computerized sub-tests within the Viennese Self-Assessment Psychology; thus, testees completed WITE-Psychology after about 50 minutes of work. The completion of WITE-Psychology took about 18 minutes. The majority of the testees (317; 78%) received a personal account to access the assessment via internet. They were advised to find a quiet place to work without disturbances and they were informed about the system requirements. The others (90, 22%) completed WITE-Psychology at the Centre of Testing and Consulting of the Viennese Faculty of Psychology. Their testing appointments were scheduled as part of a psychological counselling programme to determine their aptitude for studying psychology at the Viennese Faculty. After detecting different item functioning in the data set for the gap-text items (“ability” score), a final set of 20 items was analysed.

Results

Results of the Rasch model analyses of the three scoring variants (Hypothesis 1)

Tendency Score: Grouping two sub-samples into low-scorers and high-scorers according to the “tendency” score-split analysis resulted in a non-significant χ^2 value (Andersen $\chi^2 = 28.64$; $df = 19$, $\chi^2_{(\alpha = 1\%)} = 36.19$); the same was true when grouping sub-samples by score of the gap-text items (Andersen $\chi^2 = 26.23$; $df = 19$, $\chi^2_{(\alpha = 1\%)} = 36.19$; see Figure 2). Table 2 provides the “overestimation susceptibility” parameters of the “tendency” score ranging from -2.25 to 2.56 ($Md = -.14$, $SD = 1.26$). Figure 2 shows graphical model checks for both split criteria. Since both LRTs were not significant, there is not any further reference to them.

Alternative Tendency Score: Item 10 (experiment) with $\sigma = 1.70$ (an easy item according to the “ability” score) and item 14 (hormone) with $\sigma = -1.26$ (a rather difficult item according to the “ability” score) were excluded due to the fact that all examinees either thought they would know it or actually solved it. To run the estimation algorithm, *eRm* requires excluding all cases whose response patterns do not contribute any information. Consequently, the sample was diminished to a sample size of 345 cases. Grouping the remaining subjects in two sub-samples by score split, the analysis resulted in a non-significant χ^2 value (Andersen $\chi^2 = 14.56$; $df = 15^3$, $\chi^2_{(\alpha = 1\%)} = 30.58$; see Figure 2). The same was true when grouping sub-samples by score for the gap-text items (Andersen $\chi^2 = 24.85$; $df = 15$, $\chi^2_{(\alpha = 1\%)} = 30.58$; see Figure 2). Table 2 provides the “overestimation susceptibility” parameters of the “alternative tendency” score ranging from -2.53 to 4.75 ($Md = -.20$, $SD = 2.06$).

³ Item 5 (hypothesis) and item 19 (chance) were not well conditioned since the data structure did not allow unique estimates

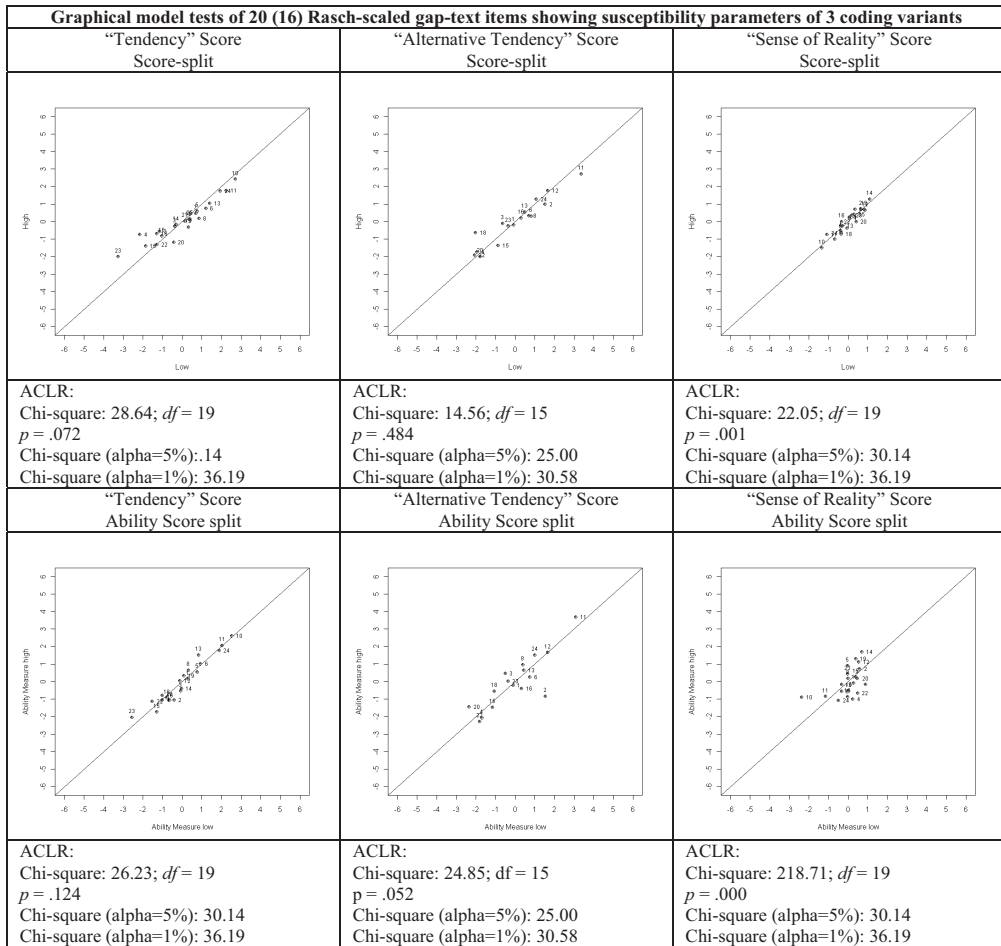


Figure 2:
Graphical model tests of 20 (16) a posteriori Rasch-scaled gap-text items

Sense of Reality Score: Grouping two sub-samples by score split (internal criterion: high vs. low “sense of reality” score), the analysis resulted in a non-significant χ^2 value (Andersen $\chi^2 = 22.05$; $df = 19$, $\chi^2_{(\alpha = 1\%)} = 36.19$; see Figure 2). When grouping sub-samples by score for the gap-text items, significant results occurred (Andersen $\chi^2 = 218.71$; $df = 19$, $\chi^2_{(\alpha = 1\%)} = 36.19$; see Figure 2). Table 2 provides the “misestimation susceptibility” parameters of the “sense of reality” score ranging from -1.41 to 1.12 ($Md = .10$, $SD = 0.64$).

Table 2:
Testing the Rasch model. Items are sorted by ascending easiness for gap-text items, higher values reflecting easier items

Items	Easiness parameter	Susceptibility parameter ("Tendency" score)	Susceptibility parameter ("Alternative tendency" score)	Susceptibility parameter ("Sense of reality" score)
24	Pavlovian conditioning	-2.179	1.812	0.557
13	Supervision	-1.754	1.204	-0.070
20	Leptosome type	-1.324	-0.868	-2.527
10	Experiment	-1.259	2.561	not analysed
6	Defense mechanism	-1.133	0.960	0.011
11	Mutation	-0.940	2.034	2.612
1	Halo-effect	-0.721	0.202	-0.721
8	Neurone	-0.531	0.498	-0.038
22	Proactive memory inhibition	-0.531	-1.312	-2.527
4	Retroactive memory inhibition	-0.365	-1.050	-2.342
3	Operant conditioning	-0.210	-0.323	-0.872
18	Oblivion curve sensu Ebbinghaus	0.449	-0.868	-1.477
5	Hypothesis	0.694	0.645	3.849
15	Self-fulfilling prophecy	0.803	-1.527	-1.792
12	Amnesia	0.933	-0.039	1.091
16	Psychoanalytic approach	1.134	-0.902	-0.335
19	Chance	1.269	0.202	4.747
2	Mediation	1.701	-0.739	0.710
14	Hormone	1.701	-0.236	not analysed
23	EEG	2.264	-2.253	-0.877
	ACLR: Chi-square: 29.83; $df=19$ $p=.054$ Chi-square (alpha=1%): 36.19	ACLR: Chi-square: 28.64; $df=19$ $p=.072$ Chi-square (alpha=1%): 36.19	ACLR: Chi-square: 14.56; $df=15$ $p=.484$ Chi-square (alpha=1%): 30.58	ACLR: Chi-square: 22.05; $df=19$ $p=.001$ Chi-square (alpha=1%): 36.19

Interpretation of the results of Rasch model analyses (Hypothesis 1)

According to the internal criterion (high vs. low “tendency” score), as well as the criterion high vs. low “ability” score, LRTs were not significant, indicating that all 20 items show a good model fit when the scoring variant “tendency” score is used. The items differentiate well between differing tendencies to overestimate, since the range of item difficulties is adequately broad. The LRT according to the internal criterion (high vs. low “alternative tendency” score) was not significant, indicating that all (here: 18) analysed items show a good model fit; the same was true for the criterion high vs. low “ability” score. The differentiation ability of the items in the “alternative tendency” score is satisfying, as the range of item difficulties is adequately broad. The LRT according to the internal criterion (high vs. low “sense of reality” score) was not significant, indicating that all 20 analysed items show a good model fit. However, the Rasch model does not hold according to the gap-text criterion (high vs. low “ability” score). The differentiation ability of the items in the alternative tendency score (of overestimation) is inadequate, as there is only a small range; this indicates that the items do not adequately differentiate between testees with differing levels of “misestimation”.

Results of the correlation coefficients of the three scoring variants with the “ability” score (Hypothesis 2)

Pearson’s correlation coefficient between the person parameter of the gap-text items and the person parameter of the overestimation items (“tendency” score) is $r = -.562$ ($n = 407$; $p < .01$). The person parameter of the gap-text items and the person parameter of the overestimation items (“alternative tendency” score) correlate with $r = .382$ ($n = 345$; $p < .01$). The correlation coefficient between the person parameter of the gap-text items and the person parameter of the misestimation items (“sense of reality” score) is $r = .343$ ($n = 407$; $p < .01$).

Interpretation of the correlation coefficients (Hypothesis 2)

The correlation between person parameter of the gap-text items (“ability” score) and the person parameter of the overestimation items (“tendency” score) is negative and small. The two scores share about 32% of variance, which means that one cannot interpret them as being independent of one another. There is hardly any correlation between person parameter of the gap-text items (“ability” score) and the person parameter of the overestimation items (“alternative tendency” score). These two scores share about 15% variance, indicating that they can be interpreted as independent of one another. The correlation between person parameter of the gap-text items (ability measure) and the person parameter of the misestimation items (“sense of reality” measure) is also negligible. The two scores share about 12% variance, indicating that they can be interpreted as independent of one another.

Discussion

Self-concept of one's ability is an important construct in the context of academic performance (Elliot & Dweck, 2005). The presented study deals with self-concept of one's domain-specific prior knowledge as one facet of self-concept of one's general ability. It was measured by calculating three different scores derived from the WITE-Psychology that are to reflect relevant aspects of the construct.

As has been shown with other item pools (Wagner-Menghin, 2004) and was hypothesized in the current study (Hypothesis 1), the present 20 items, scored using the proven scoring variant "tendency" score, fit the Rasch model; this shows that the present item pool covers the latent variable quite well. As previously shown, the "tendency" score and "ability" score derived from a sample working on a conventionally administered test are negatively correlated ($r = -.562$). This moderate negative correlation is unfortunate because theoretically the relationship between self-concept of one's ability and tested ability should be slightly positive (Furnham, 2001; Rammstedt & Ramsayer, 2002). In the past, this problem was dealt with by using an adaptive algorithm to select items to keep the number of items solved and not solved more or less constant for each person.

This problematic aspect of connectivity is overcome with the newly proposed "alternative tendency" scoring. The number of items declared to be known but not solved still serves as a measure, but missing values are assigned to solved items in order to estimate item parameters as well as person parameters; this overcomes the inherent connectivity. The items fit the Rasch model. This scoring variant facilitates the implementation of conventional fixed-test forms of LEWITE, since the new "alternative tendency" score can serve as a measure of self-concept of one's ability. Still, when using a conventional test form, one has to be aware that for persons solving all or almost all gap-text items, the "alternative tendency" measure cannot be computed. Especially problematic is that precision will be low if the tendency measure is computed using only a few items. From a practical perspective, it may be a disadvantage that calculating scores using different item samples per person requires computerized scoring software.

As mentioned above, the "alternative tendency" does not reflect underestimation of one's ability; the "tendency" score does not reflect it, either. The aspect of underestimation was implemented within the suggested "sense of reality" score. However, the conclusions are rather discouraging. Although the LRT using the internal criterion was not significant, the range of item parameters is rather limited (approximately 2 units of the logit scale are covered), pointing to problems in covering the latent continuum. A significant LRT using the external criterion high vs. low "ability" score also indicates a problem with items' function along the latent trait. Rather, one has to conclude that with this item material, a "sense of reality" for one's own verbal ability cannot be measured. The integration of tendencies of under- and of overestimation in one score did not succeed. Consequently, the negligible correlation (12% shared variance) between the "sense of reality" score and the "ability" score is of subordinate interest.

A shortcoming of the current study is that it has not explored the reliability of the "alternative tendency" measure empirically. Furthermore, the functioning of the "tendency"

score and the “alternative tendency” score using a sample taking an adaptive set of LEWITE has not yet been compared. However, future studies will address these topics.

For practical use, the “alternative tendency” score seems to be a good alternative to the “tendency” score when a conventional fixed-test of LEWITE is administered. However, considerations regarding reliability of individuals’ scores should be taken into account when interpreting individual results. As a rule of the thumb, based on Linacre’s general reliability assumptions (1993), one should refrain from interpreting the measure when the number of unsolved gap-text-items is less than 7.

Acknowledgement

We would like to express our sincere gratitude to Professor Dr. Klaus Kubinger who suggested the scoring of the “alternative tendency” score and supported this work by bringing in his ideas and expertise.

References

- Ackerman, P. L., & Wolman, S. D. (2007). Determinants and validity of self-estimates of abilities and self-concept measures. *Journal of Experimental Psychology: Applied*, 13(2), 57-78.
- Andersen, E. B. (1973). A goodness of fit test for the Rasch model. *Psychometrika*, 38(1), 123-140.
- Cattell, R. B. (1958). A need for alertness to multivariate experimental findings in integrative surveys. *Psychological Bulletin*, 55, 253-256.
- Dweck, C. S., & Leggett E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review*, (95), 256-273.
- Elliot, A. J., & Dweck, C. S. (Eds.) (2005). *Handbook of competence and motivation*. New York: Guilford Press.
- Frebort, M., & Kubinger, K. D. (2006). Ermittlung eines Anforderungsprofils für Studierende der Psychologie [Formulating requirements for Psychology students]. In B. Gula, R. Alexandrowicz, S. Strauß, E. Brunner, B. Jenull-Schiefer & O. Vitouch (Eds.), *Perspektiven psychologischer Forschung in Österreich. Proceedings zur 7. Wissenschaftlichen Tagung der Österreichischen Gesellschaft für Psychologie* (pp. 408-414). Lengerich: Pabst.
- Frebort, M., & Kubinger, K. D. (2008). Qualitätsansprüche an ein Self-Assessment zur Studienwahlberatung: Der Wiener Ansatz [Standards for self-assessments in study counselling]. In B. Hell & H. Schuler (Eds.), *Studierendenauswahl und Studienentscheidung* (pp. 95-101). Göttingen: Hogrefe.
- Furnham, A. (2001). Self-estimates of intelligence: culture and gender difference in self and other estimates of both general (g) and multiple intelligences. *Personality and Individual Differences*, 31, 1381-1405.

- Furnham, A., & Dissou, G. (2007). The relationship between self-estimated and test-derived scores of personality and intelligence. *Journal of Individual Differences*, 28(1), 37-44.
- Hong, Y.-Y., Chiu, C.-Y., Dweck, C. S., Lin, D. M. S., & Wan, W. (1999). Implicit Theories, Attributions, and Coping: A Meaning System Approach. *Journal of Personality and Social Psychology*, 77(3), 588-599.
- Kanfer, R., & Ackerman, P. L. (2005). Work Competence. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of Competence and Motivation* (pp. 336-353). New York: Guilford Press.
- Kubinger, K. D. (2009). The technique of objective personality-tests sensu R. B. Cattell nowadays: The Viennese pool of computerized tests aimed at experiment-based assessment of behavior. *Acta Psychologica Sinica*, 41(10), 1024-1036.
- Kubinger, K. D., & Litzenberger, M. (2003). Zur Validität der Objektiven Persönlichkeits-Test-Batterie "Arbeitshaltungen" [Toward the objective personality test „Arbeitshaltungen“]. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 24(2), 119-133.
- Kubinger, K. D., Moosbrugger, H., Frebort, M., Jonkisz, E., & Reiß, S. (2007). Die Bedeutung von Self-Assessments für die Studienplatzbewerbung [The meaning of self-assessments for university admission]. *Report Psychologie*, 32, 322-332.
- Linacre, J. M. (1993). Rasch-based Generalizability Theory. *Rasch Measurement Transactions*, 7(1), 283-284.
- Litzenberger, M. & Haiden, D. (2006). Evaluation einer sequentiellen Auswahlstrategie mittels Objektiver Persönlichkeitstests und Assessment Center [Evaluation of a sequential selection strategy using objective personality tests]. In T. M. Ortner, R. T. Proyer & K. D. Kubinger (Eds.), *Theorie und Praxis Objektiver Persönlichkeitstests* (pp. 234-243). Huber: Bern.
- Mair, P., & Hatzinger, R. (2006). eRm: Extended Rasch modeling. R package version 0.9.5. Retrieved April 20, 2010, from <http://r-forge.r-project.org/>.
- Poinstingl, H., Mair, P., & Hatzinger, R. (2007). *Manual zum Softwarepackage eRM (extended Rasch modeling)* [Manual for the softwarepackage eRM (extended Rasch modeling)]: *Anwendung des Rasch-Modells (I-PL-Modell)*. Lengerich: Pabst Science Publ, from <http://d-nb.info/987218735/04> / http://deposit.d-nb.de/cgi-bin/dokserv?id=3060141&prov=M&dok_var=1&dok_ext=htm.
- Rammstedt, B., & Rammsayer, T. H. (2002). Die Erfassung von selbst eingeschätzter Intelligenz. Konstruktion, teststatistische Überprüfung und erste Ergebnisse des ISI [A measure for self-estimated intelligence. Construction, psychometric proving and first results]. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 23(4), 435-446.
- Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests*. (Copenhagen, Danish Institute for Educational Research), expanded edition (1980) with foreword and afterword by B.D. Wright. Chicago: The University of Chicago Press.
- Rost, D. H., Sparfeldt, J. R., & Schilling, S. R. (2007). *Differentielles schulisches Selbstkonzept-Gitter mit Skala zur Erfassung des Selbstkonzepts schulischer Leistungen und Fähigkeiten (DISK-Gitter mit SKSLF-8)* [Differential academic self-concept grid including a scale for measuring the self-concept of academic performance and capabilities]. Göttingen: Hogrefe.

- Schmidt, L. R. (1975). *Objektive Persönlichkeitsmessung in diagnostischer und klinischer Psychologie* [Objective personality assessment in psychodiagnostics and clinical psychology]. Weinheim: Beltz.
- Sonnleitner, P., Kubinger, K. D., & Frebort, M. (2009). Das Wiener Self-Assessment Psychologie mit seinen Verfahren der experimentalpsychologischen Verhaltensdiagnostik zur Messung studienfachübergreifender Soft Skills [The Viennese Self-Assessment Psychology using experiment-based behaviour assessments for measuring study relevant soft skills]. In G. Rudinger & K. Hörsch (Eds.), *Self-Assessment an Hochschulen: Von der Studienwahl zur Profilbildung* (pp. 63-72). Göttingen: V&R unipress.
- Wagner-Menghin, M. M. (2004). *Der Lexikon-Wissen-Test (LEWITE). Computergestützte Testvorgabe und Auswertung*. [The LEWITE-Test. Computerized test administration and scoring]. In Wiener Testsystem. Mödling, Austria: Schuhfried.
- Wagner-Menghin, M. M. (2007). Conception and Construction of a Rasch-Scaled Measure for Self-Confidence in One's Vocabulary Ability. *Journal of Applied Measurement*, 8(1), 35-47.