



Psychometric properties of the Epistemological Development in Teaching Learning Questionnaire (EDTLQ): An inventory to measure higher order epistemological development

Sofia Kjellström^{1a}, Hudson Golino^b, Rebecca Hamer^c, Erik Jan Van Rossum^d, Ellen Almers^a

^aJönköping University, Jönköping, Sweden

^bUniversidade Salgado de Oliveira, Rio de Janeiro, Brazil

^cInternational Baccalaureate, The Hague, Netherlands

^dUniversity of Twente, Enschede, Netherlands

Article received 13 February / revised 5 April / accepted 25 September / available online 15 December

Abstract

Qualitative research supports a developmental dimension in views on teaching and learning, but there are currently no quantitative tools to measure the full range of this development. To address this, we developed the Epistemological Development in Teaching and Learning Questionnaire (EDTLQ). In the current study the psychometric properties of the EDTLQ were examined using a sample (N= 232) of teachers from a Swedish University. A confirmatory factor and Rasch analysis showed that the items of the EDTLQ form a unidimensional scale, implying a single latent variable (eg epistemological development). Item and person separation reliability showed satisfactory levels of fit indicating that the response alternatives differentiate appropriately. Endorsement of the statements reflected the preferred constructivist learning-teaching environment of the response group. The EDTLQ is innovative since is the first quantitative survey to measure unidimensional epistemological development and it has a potential to be used as an apt tool for teachers to monitor the development of students as well as to offer professional development opportunities to the teachers.

Keywords: Rasch modelling, unidimensional development, epistemological development, good teaching, application, good textbook

¹ Contact information: Sofia Kjellström, The Jönköping Academy for Improvements of Health and Welfare, School of Health & Welfare, Jönköping University, P.O. 1056, S-55111 Jönköping, Sweden. E-mail: sofia.kjellstrom@ju.se. DOI: <http://dx.doi.org/10.14786/flr.v4i5.239>



1. Introduction

In the post-industrial society, global issues such as climate change, sustainable economic development and increasing wellbeing worldwide have highlighted the need for education that promotes the ability to think critically and for problem solvers who can use conflicting evidence to understand and address complex or so-called *wicked* problems. Such problem solvers need to be able to tolerate ambiguity, take different perspectives and acquire new knowledge throughout life, skills associated with higher order thinking and complex meaning making (e.g. Kegan, 1994; Van Rossum & Hamer, 2010), developed epistemic thinking (e.g. Barzilai & Eshet-Alkalai, 2015), and flexible performance associated with deep understanding (Perkins, 1993). The need for more of such problem solvers underlies the current focus on characteristics and skills currently referred to as 21st century skills which include “communication skills, creativity, critical reflection (and self-management), thinking skills (and reasoning), information processing, leadership, lifelong learning, problem solving, social responsibility (ethics and responsibility) and teamwork” (Strijbos, Engels & Struyven, 2015, p20). Although one can safely argue these skills are not unique or particularly novel to the 21st century (e.g. Mishra & Kereluik, 2011; Voogt et al., 2013), the consensus is that the complexity and global nature of the major problems that confront humanity now and in the future increased the need for this type of thinking and problem solving.

Formal education, and higher education (HE) in particular, is an essential tool to ensure a society will include psychologically mature citizens (e.g. Piaget, 1954; Kohlberg, 1984; Schommer, 1998) with a well-developed ability to think critically. For this, it needs to develop students’ understanding of how knowledge is created, its scope and what constitutes justified beliefs and opinions: in short it needs to focus on ensuring students’ epistemological development. However, currently many HE graduates have not developed towards the lowest level of critical thinking and reflective judgment in the sense of routinely using established procedures and assumptions within one discipline or system to evaluate knowledge claims and solve ill-structured problems well (Kegan, 1994; King & Kitchener, 2004; Van Rossum & Hamer, 2010; Arum & Roksa, 2011), let alone that such graduates have developed a theory of knowledge that accommodates differences in methods and procedures across systems and disciplines that would be necessary to address the global issues and wicked problems facing us. Indeed, a range of studies indicate that education focusing on reproduction of established knowledge and procedures can lead to poorer learning strategies (e.g. Newble & Clarke, 1986; Gow & Kember, 1990; Linder, 1992) and occasionally to regression in epistemic beliefs (Perry, 1970) and decreased student well-being (e.g. Van Rossum & Hamer, 2010; Lindblom-Ylläne & Lonka, 1999; Yerrick, Pedersen, & Arnason, 1998). To increase education’s success rate, curricula at all age levels would need to address and assess development in students’ epistemic beliefs, choosing instruction methods that emphasize the complexity of understanding, that stimulate students’ development of value systems and epistemic reflection. Whilst literature points towards a relationship between teachers’ epistemological development and identity development, and their success in encouraging similar development in their students (see for a review Van Rossum & Hamer, 2010), as yet it is not easy to establish which teachers are more successful in modelling complex meaning making and so in teaching students complex value systems.

With regard to the further life span, research on adult development shows that epistemological development does not need to end after adolescence or leaving formal education. Adults may develop increasingly complex meaning and value systems (Commons, 1989, 1990; Demick & Andreoletti, 2003; Fischer & Pruyne, 2003; Hoare, 2006; Loevinger & Blasi, 1976), with convincing evidence that the later phases of personality development lead to higher levels of critical and meta systemic thinking (i.e. the ability to see how different systems interact with each other), responsibility, positive valuation of human rights, extended time horizon and perspective consciousness (Commons & Ross, 2008a, 2008b; Kjellström & Ross, 2011; Sjölander, Lindström, Eriksson & Kjellström, 2014; Kjellström & Sjölander et al, 2014).

A wealth of research exists describing students’ and teachers’ views on the nature and source of knowledge, measured both quantitatively and qualitatively, proposing a unidimensional model of epistemological sophistication (e.g. Perry, 1970; Baxter Magolda, 1992, 2001; Kegan, 1994; Kuhn,



1991; see also Van Rossum & Hamer, 2010 for a review linking qualitative and quantitative models). Student and teacher views on the nature of knowledge have been linked to views on learning, good teaching, understanding and application (Van Rossum & Hamer, 2010). The first major problem is that the existing measurements in this field are time consuming both for students (participants) and teachers (or researchers). The students need to be interviewed or write long essays on dilemmas. Secondly, the researchers' analysis procedures are time consuming and/or require scoring skills that need to be trained and often are only honed over time. Thirdly, many of these models or qualitative empirical studies have not been designed with a large enough sample size to capture the relatively rare sophisticated epistemic perspectives and belief structures (Kegan, 1994; Van Rossum & Hamer, 2010), for instance the recent research regarding conceptions of understanding (e.g. Irving & Sayre, 2013) has not progressed beyond the flexible performance interpretation championed by Perkins (1993) in *Teaching for understanding*, whilst at least two more complex interpretations of understanding have now been empirically observed (Hamer & Van Rossum, 2016). Existing quantitative measures, such as the EBQ (Schommer, 1994), ILS (Vermunt, 1996) or the dilemma or scenario based assessment tool explored by Barzilai and Weinstock (2015), on the other hand, also do not include the more sophisticated views on learning and knowledge (see for an extensive discussion of the ILS and EBQ Van Rossum & Hamer, 2010, chapter 4).

Therefore, a questionnaire measuring the more sophisticated views on teaching and learning concepts would be a valuable tool for teachers to monitor the effectiveness of a curriculum designed to encourage students' cognitive identity development, complex meaning making and their adoption of a complex value system. Similarly, given that the more complex epistemological approach the teacher has, the more likely it is that students develop similar capabilities (Van Rossum & Hamer, 2010). Such a tool would be useful to educational management to know when to offer professional development opportunities to teachers, thereby ensuring that their staff is indeed able to shape the developmental educational environment necessary to support students' epistemological development to the higher levels of complex thinking. However, as many of the current models and data collection tools do not include these higher levels of epistemic thinking, to construct such a tool requires a review of models that have sufficient evidence and examples of student, teacher and adult thinking reflecting these more elusive levels of complex thinking.

In preparation towards developing such a questionnaire a review of existing models was undertaken, merging evidence of student and teacher thinking with that regarding adult development models (e.g. Baxter Magolda, 1992, 2001; Belenky, Clinchy, Goldberger & Tarule, 1986; Dawson, 2006; Kegan, 1994; King & Kitchener, 1994; Kuhn, 1991; Perry, 1970; Van Rossum & Hamer, 2010; West, 2004). From this review, the six stage developmental model of learning-teaching conceptions developed by Van Rossum and Hamer (2010) and expanded upon in a number of follow up studies (Van Rossum & Hamer, 2012, 2013; Hamer & Van Rossum, 2016) was selected as the primary source for constructing items and scales designed to access a broader range of epistemological development and through this psychological maturity within a teaching and learning environment (see Table 1). An important consideration here was that as of 2016, this model is supported by the narratives of more than 1,200 students, including repeated measurements and longitudinal student data evidencing the developmental aspect of the model. Further, although it only covers approx. 70 teacher narratives, the model is proposed to model both student and teacher thinking (Richardson, 2012) and would therefore also model adult thinking to some degree at least.

The result is the *Epistemological Development in Teaching and Learning Questionnaire* (EDTLQ) that aims to measure stages of development in the domain of teaching and learning. A number of items (statements) was constructed, covering five issues/domains in learning and teaching. Domains included views on 'good' teaching, understanding, application, good classroom discussion (Van Rossum & Hamer, 2010; Hamer & Van Rossum, 2016) and a good textbook (Van Rossum & Hamer, 2013). A sixth scale was constructed based upon developmental responsibility research (Kjellström, 2005; Kjellström & Ross, 2011) and unpublished preliminary results from an ongoing longitudinal study on responsibility for learning. The items were selected to represent a sequence of learning and teaching conceptions, ranging from level 2 (L2, Memorizing) to level 6 (L6, Growing



self-awareness) (Van Rossum & Hamer, 2010, see and Table 1), which the authors believed should appeal in different ways to various teachers and students as appropriate to their level of thinking within the domain of epistemological development (see Table 2). The different scales developed here refer to specific and shared experiences in the teaching-learning environment, and so aim to address a weakness in other measurement tools regarding “the use of abstract and ambiguous words [or] items which are too general or vague to allow for a consistent point of reference” (Barzilai & Weinstock, 2015, p. 143). In the questionnaire items from each scale were presented in a random order. The Swedish and English version were created in a continuous process, were the items were translated back and forth several times by native speakers among the current authors.

The construction of different scales representing different aspects of students’ views on the learning-teaching environment may be perceived to point towards a multi-dimensional interpretation regarding students’ epistemic beliefs and epistemological development. Indeed traditionally, in developing a tool to measure differences in epistemic beliefs, respondents would be asked to indicate the extent to which each item reflected their beliefs. Factor analysis of the responses ideally would then identify which sets of items are strongly correlated, i.e. load significantly on one or another factor. Each factor then would be identified as representing a learning-teaching conception or level of development. This approach underlies existing tools, such as the ILS (Vermunt, 1996) and the more recently approach to measure epistemic thinking by Barzilai and Weinstock (2015). In the EBQ originally developed by Schommer (1990, 1994) and expanded upon by a range of followers (e.g. Qian & Alverman, 2000; Qian & Pan, 2002; Paulsen & Feldman, 2005; Bråten & Strømsø, 2004; Schreiber & Shinn 2003), the factors refer to (possibly orthogonal) trajectories of epistemic development. The scales developed in the current study are however proposed to represent different contexts in which a student’s underlying epistemic belief system will express itself. The items present descriptions of aspects of different epistemological ecologies (Van Rossum & Hamer, 2010): profiles or constellations of beliefs that are closely linked and which change more or less simultaneously when a respondent’s perspective shifts to a more complex epistemic level. In this sense, Van Rossum and Hamer’s 2010 model is one of the family of models that “suggest that there is ... an underlying trajectory explaining development across domains and topics [i.e.] ... a more global development” (Barzilai & Weinstock, 2015, p. 144), an assumption that is often disputed. The aim of this study is to explore the assumption of an underlying unidimensional development trajectory or “more global development” (Barzilai & Weinstock, 2015).

Traditionally, respondents are asked to endorse each item as it is presented if it is a part of their epistemic position, i.e. how they view (their own) learning or experience teaching. In that case, the hierarchical inclusiveness of the developmental model underlying the item construction implies that items representing less sophisticated views on learning would be endorsed by more respondents, whilst at least in theory, the items reflecting the more sophisticated levels are expected to be endorsed by fewer respondents. In this sense the items reflecting the more sophisticated levels would seem ‘more difficult’ to endorse. However, participants completing the items of the EDTLQ were instructed to rate the statements in accordance with how important the statements were to teaching and learning on a five-point ordinal scale, ranging from least important (1) to most important (5). By rating the items to the respondents’ perception of most to least importance to learning and teaching, the qualitative empirical evidence used by Van Rossum and Hamer (2010; Hamer & Van Rossum, 2016) points towards a different response pattern that fits Kegan’s consistency hypothesis. Kegan (1994) states that people strongly prefer to function at their highest level of epistemic thinking and when circumstances prevent the expression of this highest level they feel unhappy. The consistency hypothesis then would predict that respondents will prefer items, i.e. rate as most important or find easier to endorse, those that reflect their own current epistemic beliefs, and will reject all others. The items reflecting the lower levels of development, which in a traditional approach would be relatively easy to endorse, will now be rejected – i.e. difficult to endorse – because they reflect a view on learning-teaching that the respondent has moved beyond. The items that reflect thinking that is more sophisticated than that of the respondent are rejected, as they do not reflect a current aspect of the respondent’s thinking about learning, teaching and knowing. This means that the choice to ask



respondents to rate the items by importance will lead to an answer pattern that will reflect the level of development, a learning-teaching conception, of the majority of the respondent group.

The empirical qualitative evidence of over 1200 HE students and 70 HE teachers (Hamer & Van Rossum, 2016) suggests that a respondent group of freshmen or sophomore HE students will result in the endorsement of items on levels 2, 3 in a traditional learning environment, or in a more constructivist teaching-learning environment a preference for 3 and 4 level items, with items reflecting level 1, 5 and 6 being rejected by most respondents, i.e. being most difficult to endorse. Again depending on the level of traditional or constructivist teaching practices, a sample of teachers will either demonstrate a similar pattern to students (in predominantly traditional teaching environments or secondary school level), whilst in a predominantly constructivist teaching environment respondents would find items at levels 3, 4 and perhaps 5 to be most easy to endorse, rejecting those reflecting level 1, 2 and 6. For teachers however there is a complicating factor to interpreting their responses. Teacher self-generated responses regarding good teaching or their expectations regarding student learning are often seen to include large quantities of academic discourse, occasionally to an extent that it obscures the formulation of their own views on teaching or learning (Säljö, 1994, 1997), implying that their self-generated responses (e.g. in interviews or narratives) may reflect social desirable answering patterns to a greater degree than that of students. It is unknown to what extent similar teacher response bias or patterns may express themselves in the EDTLQ.



Table 1

Epistemological profiles within Van Rossum & Hamer six stage developmental model of learning-teaching conceptions

| | Van Rossum & Hamer (2010) | | | Hamer & Van Rossum (2010) | | Hamer & Van Rossum (2016) | | Van Rossum & Hamer (2013) |
|----|---|--|--|--|---|---|---|--|
| | Learning conception | Teaching conception | Application | Core learning question | Student-teacher relationship | Understanding | Assessment of understanding | Good text book |
| L1 | Increasing knowledge | Imparting clear/well structured knowledge | Comparing facts to reality | - | Obedience | Understanding every word, every sentence | Knowledge questions | - |
| L2 | Memorising | Transmitting structured knowledge (acknowledging receiver) | Reproducing at exams | What do I need to know (for the exam)? | I may ask questions | Answering exam questions, reproduction of (certain) knowledge | Multiple choice (recognising) Open questions (reproduction) | Current language, clear chapter structure to help recall: ie a preview, keywords in margin, summary and test questions |
| L3 | Reproductive understanding/ application or Application foreseen | Interacting and Shaping | Answering exam questions; Using knowledge algorithmically in practice | What will I need later (in life or my career)? | The expert (teacher) doesn't know everything either, I can have my own opinion and all opinions are equally valid and should be respected | Applying knowledge to practical situations (simple) | Frequent testing (parts of total) Open questions (reproduction) Practical cases and assignments (practical application and own ideas) | Good chapter structure, includes examples from practice and applications that help to understand and memorise |



| | | | | | | | | |
|----|------------------------------|--|---|--|--|---|---|---|
| L4 | Understanding subject matter | Challenging to think for yourself / developing a way of thinking | Using knowledge in flexible ways, within and outside the educational setting; 'creativity' is reached by following disciplinary rules | How do I formulate a good argument, how to approach a complex or ill structured problem? | A teacher should help me learn to think for myself. Opinions and conclusions need to be evidence based | Reorganising knowledge into your own systematic structure. | Open questions (independent thinking, analytical skills) Interesting cases and assignments (application in unfamiliar situations) Questions and cases need to be interesting and motivating | A book that makes you think, content increases in complexity and depth, examples and images; simplicity in presentation |
| L5 | Widening horizons | Dialogue teaching | Problem solving in a heuristic and relativist way | What does it look like if I take a different perspective? | Equality between student and teacher. | Formulating arguments for or against, and using what is learned in your own argumentations, multiple perspectives | Assignments for assessing taking different/other perspective(s) Oral / interview / debate | Every page leads to more questions, not one truth/ perspective, makes you see issues in a different light |
| L6 | Growing self awareness | Mutual trust and authentic relationships: Caring | - | What fits me; what perspective, which choices show who I am? | The teacher is subservient to my growth as a person | A basis to build your life around, a temporary relief from uncertainty, less stagnation in living; getting closer to the self | Life for assessing self-knowledge and the development of self-awareness: 'making your own way' | Consists of self-chosen materials, follows one's own interests |



Table 2

EDTLQ items, codes and scoring levels according to Van Rossum & Hamer (2010) and Hamer & Van Rossum (2016)

| Teaching and Learning | code | Scoring level |
|--|------|---------------|
| A good study book | | |
| 1. is well structured, gives the main points and the consensus story of the subject. | sb1 | L2 |
| 2. provides many up-to-date examples and examples from practice. | sb2 | L3 |
| 3. presents alternative conclusions, shows different interpretations within a subject matter or clarifies different perspectives on the content. | sb3 | L5 |
| 4. evokes a critical attitude and invites reflection. | sb4 | L4 |
| 5. provokes and challenges the students' pre-conceptions. | sb5 | L4 |
| 6. makes students think about the fundamental question of knowledge: "How can we know this?" | sb6 | L5 |
| Discussions during a course are good when | | |
| 1. the same students don't dominate. | di1 | L2 |
| 2. the teacher answers students' questions. | di2 | L2 |
| 3. different solutions are illustrated from multiple perspectives. | di3 | L5 |
| 4. students and teachers learn together and from each other. | di4 | L5 |
| 5. opinions are based on and/or supported by evidence. | di5 | L4 |
| 6. you hear all the different views that people have. | di6 | L3 |
| Being able to apply knowledge means to | | |
| 1. pass one's exams. | ap1 | L2 |
| 2. look at a problem from multiple perspectives. | ap2 | L5 |
| 3. be able to solve familiar problems using what you have learnt. | ap3 | L3 |
| 4. be able to solve real life problems by combining knowledge and skills in new ways. | ap4 | L4 |
| 5. support your own point of view on an issue using evidence and facts. | ap5 | L4 |
| 6. use your knowledge to make the world a better place for those around you (society, human kind and/or nature). | ap6 | L6 |



To understand something means

- | | | |
|--|-----|----|
| 1. to be able to apply and use it properly. | un1 | L3 |
| 2. you know something by heart. | un2 | L1 |
| 3. to see the connections within a larger context. | un3 | L4 |
| 4. to see the underlying assumptions and how these lead to particular conclusions. | un4 | L5 |
| 5. to realize how different perspectives influence what you see and understand. | un5 | L5 |
| 6. to be able to answer test or exam questions correctly. | un6 | L2 |

For students to take responsibility for learning means that students*

- | | | |
|---|-----|----|
| 1. come to classes well prepared, take notes, review the notes afterwards, and complete assigned tasks. | re1 | L2 |
| 2. have an active interest in the course and are motivated to learn. | re2 | L3 |
| 3. take responsibility for their own personal development. | re3 | L5 |
| 4. follow their own interests in searching for knowledge. | re4 | L6 |
| 5. are aware of their strengths, monitoring their progress and making adjustments when necessary. | re5 | L4 |
| 6. systematically reflect on what they have learned and how it fits in what they already know. | re6 | L4 |

The best teaching ought to

- | | | |
|--|-----|----|
| 1. focus on conveying the most important facts and knowledge as clearly as possible. | bt1 | L2 |
| 2. inspire the students so that they are motivated to learn. | bt2 | L4 |
| 3. focus on application of what is learnt. | bt3 | L3 |
| 4. make students question their current view on the world. | bt4 | L5 |
| 5. help students realize the limitations of current knowledge and understandings. | bt5 | L5 |
| 6. make sure that the whole content of the course is covered. | bt6 | L1 |

*) these items are not empirically derived from Van Rossum and Hamer's developmental model used for the other scales. The responsibility item scale is based developmental research literature and work on responsibility issues (Kjellström, 2005; Kjellström & Ross, 2011) and unpublished preliminary results from an ongoing longitudinal study on responsibility for learning.



2. Aim

The aim of the current paper is to investigate the psychometric properties of EDTLQ via confirmatory factor analysis and the Rasch model for dichotomous responses, in order to answer the following questions:

1. Do the items of the EDTLQ present an adequate fit to a unidimensional model of epistemological development?
2. Do the items present an adequate fit to the Rasch model for polytomous responses?
3. Are the items' response categories performing as intended?
4. Do the item parameters (difficulty to endorse) follow the order predicted by theory and the consistency hypothesis?

3. Materials and Method

3.1 Design and setting

The study was designed as an online survey involving both staff and students in higher education, with the aim to develop a quantitative tool to measure psychological maturity using a variety of existing and newly developed survey items. In this study we report on the results of survey among staff, i.e. university teachers and researchers. The analysis focuses on the psychometric properties of the newly developed EDTLQ with regard to the fit of the items and scales developed.

3.2 Participants

The participants comprise a convenience sample of staff which included teachers and researchers working in one of the four schools at Jönköping University: The School of Health and Welfare, the School of Education and Communication, Jönköping International Business School and the School of Engineering. The official language is Swedish at three of the schools and English at the remaining school. About two thirds of the respondents approached opened the questionnaire (N=340, 68%) resulting in 232 full responses.



Table 3
Participant demographics

| Age | N (%) | Highest level of education completed | N (%) | Teaching experience | N (%) |
|-------------------|---------|--------------------------------------|---------|---------------------|---------|
| 29 yrs or younger | 11 (4) | 3 yrs (BA degree equivalent) | 6 (2) | None | 5 (2) |
| 30-39 yrs | 57 (21) | 4 yrs (MA degree equivalent) | 26 (10) | 1 yr or less | 14 (5) |
| 40-49 yrs | 87 (33) | 5 yrs (MA degree equivalent) | 80 (30) | 1-5 yrs | 51 (19) |
| 50-59 yrs | 67 (25) | Licentiate (2 yrs doctorate) | 14 (5) | 6-10 yrs | 63 (24) |
| 60 yrs or more | 45 (17) | PhD (4 yrs doctorate) | 78 (29) | 11-20 yrs | 86 (32) |
| | | Associate Professor | 24 (9) | 20 yrs or more | 47 (18) |
| | | Professor | 37 (14) | | |

The majority of the respondents were well established teachers: half of the respondents had 11 years or more teaching experience, and fifty-eight per cent were aged between 30 and 59 years old. Seven per cent have no, or less than a year teaching experience which could mean that they were recently employed, but all have studied at university level which means that all are familiar with education and learning concepts. As expected from university teachers and researchers, they were highly educated with almost half (48%) having achieved the Swedish licentiate, doctorate, associate or full professorship. More detailed demographics of the respondents are provided in Table 3.

3.3 Data collection

The data were collected by two of the current authors in 2014. A web-based questionnaire was sent to all respondents by e-mail. The invitation e-mail included information about the study and a booklet about the study. The questionnaire and all supporting information were provided in both English and Swedish, with respondents given the option to responding to the survey in either language. Two reminders were sent to the participants.

3.4 Research ethics

All respondents were informed that participation was entirely voluntary, of their right to withdraw at any time, and the confidential treatment of their responses during analysis and in reporting. The study was approved by Regional Ethical Review Board in Linköping.

3.5 Measures/Instruments

The full survey comprised two smaller sections: one covering socio-demographic information (i.e., gender, age, education), and the other comprising the EDTLQ. The items of the EDTLQ are presented in Table 2 above. Table 4 shows how the items are expected to cluster to represent the different levels of thinking or ways of knowing in order of complexity according to developmental theory and the model of Van Rossum and Hamer (2010, 2013; Hamer & Van Rossum, 2016).



Table 4

Items of the EDTLQ by assumed level of epistemological development and complexity of thinking (Van Rossum & Hamer, 2010, 2013; Hamer & Van Rossum, 2016)

| Level | Study book (sb) | Discussion (di) | Application (ap) | Understanding (un) | Responsibility for learning (re)* | Best teaching (bt) |
|-------|--------------------|--------------------|---------------------|-----------------------|--------------------------------------|-----------------------|
| 1 | | | | un2 | | bt6 |
| 2 | sb1 | di1, di2 | ap1 | un6 | re1 | bt1 |
| 3 | sb2 | di6 | ap3 | un1 | re2 | bt3 |
| 4 | sb4, sb5 | di5 | ap4, ap5 | un3 | re5, re6 | bt2 |
| 5 | sb3, sb6 | di3, di4 | ap2 | un4, un5 | re3 | bt4, bt5 |
| 6 | | | ap6 | | re4 | |

*) these items are not empirically derived from Van Rossum and Hamer’s developmental model used for the other scales. The responsibility item scale is based on developmental research literature and work on responsibility issues (Kjellström, 2005; Kjellström & Ross, 2011) and unpublished preliminary results from an ongoing longitudinal study on responsibility for learning.

3.6 Data analysis

The developmental literature (Commons et al., 2008a, 2008b; Dawson, Goodheart, Draney, Wilson, & Commons, 2010; Golino, Gomes, Commons & Miller, 2014) uses the Rasch family of statistical models to assess the quality of its instruments, as well as the expected equivalence and order of items (Andrich, 1988; Rasch, 1960). The benefits of using the Rasch family of models for measurement involve the construction of objective and additive scales, with equal-interval properties (Bond & Fox, 2001; Embreston & Reise, 2000). It also produces linear measures, gives estimates of precision, allows the calculation of quality of fit indexes and enables the parameters’ separation of the object being measured and of the measurement instrument (Panayides, Robinson & Tymms, 2010). Rasch modeling enables the reduction of all items of a test into a common developmental scale (Demetriou & Kyriakides, 2006), collapsing in the same latent dimension person’s abilities and item’s difficulty (Bond & Fox, 2001; Embreston & Reise, 2000; Glas, 2007), as well as enables verification of the hierarchical sequences of both item and person, which is especially relevant to developmental stage identification (Golino et al, 2014; Dawson, Xie & Wilson, 2003).

The simplest model is the dichotomous Rasch Model (Rasch, 1960/1980) for binary (i.e. one or the other) responses. It establishes that the right/wrong scored response X_{vi} , depends upon the performance β of that person and on the difficulty δ of the item. Andrich (1978) extended the classic Rasch Model for dichotomous responses to accommodate polytomous responses with the same number of categories, also called the *rating scale model (RSM)*. The rating scale model constrains the category-intersection distances (thresholds) to be equal across all items (Andrich, 1978; Mair & Hatzinger, 2007a, 2007b). In the rating scale model the item difficulty is interpreted as the resistance to endorsing a rating scale response category.

Although some authors point out that the Rasch models are the simplest forms of Item Response Theory (IRT; Hambleton, 2000), Andrich (2004) argues this is an irrelevant argument either way, as IRT and Rasch models differ in nature and epistemological approach. In IRT one chooses the model to be used (one, two or three parameters) according to which better accounts for the data, while in the Rasch paradigm the model is used because “it arises from a mathematical formalization of invariance which also turns out to be an operational criterion for fundamental measurement” (Andrich, 2004, p. 15). Instead of data modeling, the Rasch’s paradigm focuses on the verification of data fit to a fundamental measurement criterion, compatible with those found in the physical sciences (Andrich, 2004).



A confirmatory factor analysis was conducted prior to the rating scale model, using the *lavaan* package (Rosseel, 2012), in order to verify if the questionnaire presented an adequate fit to a unidimensional model. Verifying the questionnaire dimensionality is relevant given the unidimensionality assumption of the rating scale model. The estimator used was the robust weighted least squares (WLSMV). The data fit to the model was verified using the root mean-square error of approximation (RMSEA), the comparative fit index (CFI: Bentler, 1990), and the non-normed fit index (NNFI: Bentler & Bonett, 1980). A good data fit is indicated by a RMSEA shorter or equal than 0.05 (Browne & Cudeck, 1993), or a stringent upper limit of 0.07 (Hu & Bentler, 1999), a CFI equal to or greater than 0.95 (Hu & Bentler, 1999), and a NNFI greater than 0.90 (Bentler & Bonett, 1980). The reliability of the general factor was calculated using the *semTools* package (Pornprasertmanit, Miller, Schoemann, & Rosseel, 2013).

In order to apply the rating scale model, the *eRm* package (Mair, Hatzinger, & Maier, 2014) from the R software (R Core Team, 2013) was used. The R software is a free and open source system for statistical computing and graphics creation that was originally developed by Ross Ihaka and Robert Gentleman at the Department of Statistics, University of Auckland, New Zealand (Hornik, 2014). Being free and open source, makes R an attractive tool to use as it facilitates replication and verification by other researchers. Most importantly however is that it is flexible: the user can perform various analyses, implement different techniques of data processing, as well as generate graphs of multiple types on a single platform to a greater degree than in many commercial software packages.

The application of the rating scale model through the *eRm* package (Mair, Hatzinger, & Maier, 2014) followed a four step procedure described below. First, the dataset was subset in order to exclude cases with missing data in all items. Secondly, the rating scale model was applied using the *RSM* function of the *eRm* package. Then, a graphic analysis was made to detect issues in the categories. Andrich (2011) points out a large difference between the traditional IRT paradigm and the Rasch paradigm when dealing with rating scales. The former takes the ordering of the categories for granted, while in the latter the order is a hypothesis that needs to be checked (Andrich, 2011). If the categories do not follow the hypothesized order (e.g.: reversed categories $C_1 < C_3 < C_4 < C_2$), then they will need to be examined further and improved experimentally (Andrich, 2011). Also, it is necessary to identify whether each category has a probability of being chosen (or marked) greater than the other categories in a specific range of the latent continuum. Graphically, it means that C_X 's category characteristic curve cannot be contained inside any other category curve. Finally, the fourth step in the rating scale analysis involved checking the fit to the model. The information-weighted fit (Infit) mean-square statistic was used. It represents "the amount of distortion of the measurement system" (Linacre, 2002. p.1). Values between 0.5 and 1.5 logits are considered productive for measurement. Values smaller than 0.5 and between 1.5 and 2.0 logits are not productive for measurement, but do not degrade it (Wright, Linacre, Gustafson & Martin-Lof, 1994).

It is important to note that respondents were instructed to rate the item statements in accordance with how important they were to learning and teaching on a five-point ordinal scale, where 1 is least important and 5 is most important. This affects the interpretation of the item parameters: Items with high difficulty estimates (i.e. with high resistance to endorsing a rating scale response category) are those which fewer people considered important to their views and opinions, while items with low difficulty estimates (i.e. with low resistance to endorsing a rating scale response category) are those that people considered very important to their views. Thus, to reflect increasing psychological maturity the interpretation of the latent dimension must be reversed in the present study: from statements most important to a person's view (easier to endorse) to statements least important to a person's view (more difficult to endorse).

4. Results

A first analysis focused on the first research question regarding the items of the EDTLQ fitting a unidimensional model. The confirmatory factor analysis showed that the unidimensional model, with one latent variable (f1 = factor1) explaining the 36 questionnaire items, presented an adequate data fit ($\chi^2(594) =$



816, $p = 0.000$, RMSEA = 0.05, CFI = 0.95, NNFI = 0.94). The standardized factor loadings ranged from 0.16 (item un2) to 0.77 (item un4). The reliability of the general factor was 0.91 calculated using both the Cronbach's alpha and the coefficient omega (Raykov, 2001). The standardized estimates of the unidimensional model were plotted using the *semPlot* package (Epskamp, 2014).

The factor structure is represented in Figure 1 as a weighted network, where the higher the standardized factor loading the thicker and more saturated is the arrow. The items un2 ('you know something by heart'), di2 ('when the teacher answers the students' questions), un6 ('able to answer test or exam questions correctly'), ap1 ('pass one's exams') and di6 ('hear all the different views that people have') all have a factor loading smaller than 0.3 and reflect the less sophisticated levels of epistemological development (see Table 2). The items with the largest factor loadings (> 0.55) in majority reflect the more sophisticated levels of epistemological thinking of the model of Van Rossum and Hamer (2010, 2013), i.e. di3 ('different solutions are illustrated from different perspectives'), bt3 ('focus on application on what is learnt'), un5 ('realise how different perspectives influence what you see and understand'), un4 ('see the underlying assumptions and how these lead to particular conclusions') and un3 ('to see connections within a larger context').

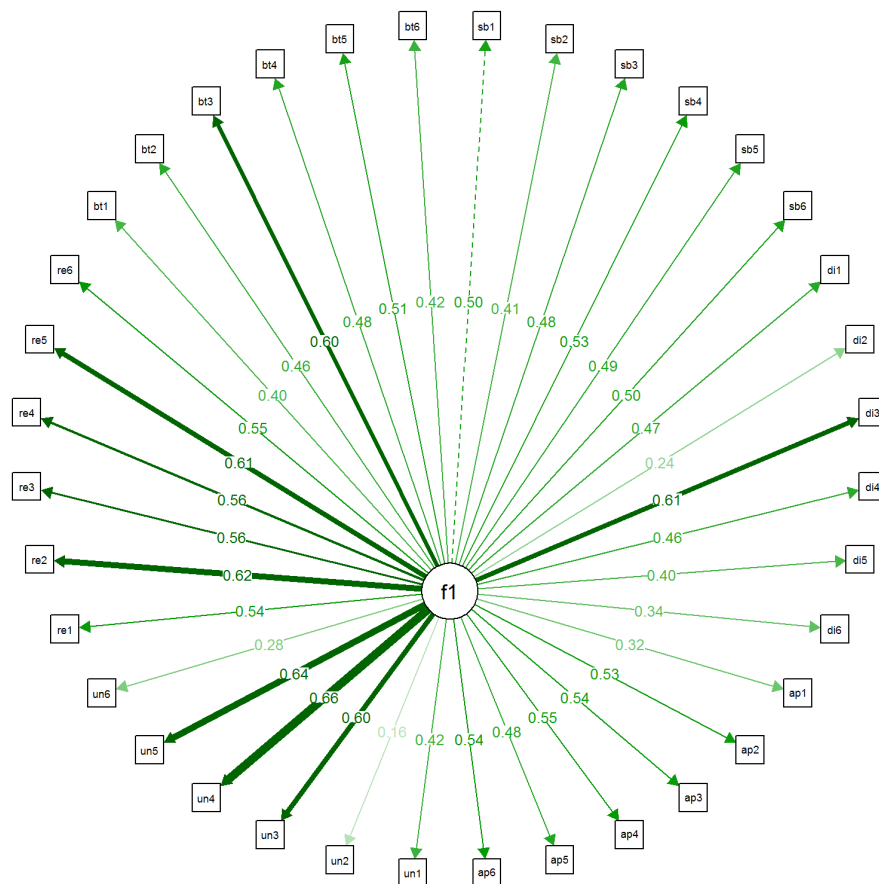


Figure 1. Unidimensional factor structure of the Teaching and Learning Questionnaire

The relevance of conducting a confirmatory factor analysis prior to the Rasch analysis lies in the assessment of the unidimensionality assumption. This is one of the available strategies, i.e. one can verify if the data fits a unidimensional model using confirmatory factor analysis and then proceed to the Rasch analysis.



The rating scale model results indicated an adequate fit of the items to the unidimensional model (see Table 5), with mean infit meansquare of 0.97 (Min = 0.69, Max = 1.30). These values are within the range considered productive for measurement (Wright, Linacre, Gustafson & Martin-Lof, 1994). The location parameter of Table 5 is basically the item difficulty (i.e. the resistance to endorsing a rating scale response category), while the threshold are the points where the category curves intersect. So, threshold 1 is the point in the latent variable where the category 1 ('least important', value = 0) intersects with category 2 (value = 1; Figure 1). This means that people with ability or proficiency in the latent variable equal to threshold 1 (-0.41) will have equal probability of choosing item sb1 to reflect category 1 or category 2.

Table 5

Item Infit, location and thresholds of items in the EDTLQ

| Item | Infit Meansquare | Location | Threshold 1 | Threshold 2 | Threshold 3 | Threshold 4 |
|------|------------------|----------|-------------|-------------|-------------|-------------|
| sb1 | 1.06 | 0.82 | -0.41 | 0.14 | 1.17 | 2.37 |
| sb2 | 0.97 | 1.44 | 0.21 | 0.76 | 1.79 | 2.99 |
| sb3 | 0.92 | 0.89 | -0.34 | 0.22 | 1.24 | 2.44 |
| sb4 | 0.80 | 0.49 | -0.74 | -0.18 | 0.84 | 2.04 |
| sb5 | 1.17 | 1.35 | 0.12 | 0.67 | 1.70 | 2.90 |
| sb6 | 1.21 | 1.29 | 0.06 | 0.62 | 1.65 | 2.85 |
| di1 | 1.03 | 1.32 | 0.09 | 0.65 | 1.68 | 2.88 |
| di2 | 1.30 | 2.48 | 1.25 | 1.81 | 2.83 | 4.03 |
| di3 | 0.82 | 0.49 | -0.74 | -0.19 | 0.84 | 2.04 |
| di4 | 1.07 | 0.82 | -0.41 | 0.15 | 1.17 | 2.38 |
| di5 | 0.89 | 0.99 | -0.24 | 0.31 | 1.34 | 2.54 |
| di6 | 1.20 | 2.29 | 1.05 | 1.61 | 2.64 | 3.84 |
| ap1 | 1.04 | 2.60 | 1.37 | 1.93 | 2.95 | 4.15 |
| ap2 | 0.86 | 0.87 | -0.36 | 0.19 | 1.22 | 2.42 |
| ap3 | 1.02 | 1.23 | 0.00 | 0.55 | 1.58 | 2.78 |
| ap4 | 1.05 | -0.65 | -1.88 | -1.32 | -0.29 | 0.91 |
| ap5 | 1.03 | 0.92 | -0.32 | 0.24 | 1.27 | 2.47 |
| ap6 | 1.07 | 1.15 | -0.08 | 0.47 | 1.50 | 2.70 |
| un1 | 1.08 | 1.10 | -0.13 | 0.42 | 1.45 | 2.65 |
| un2 | 1.23 | 3.87 | 2.64 | 3.20 | 4.22 | 5.42 |



| | | | | | | |
|-----|------|------|-------|-------|------|------|
| un3 | 0.76 | 0.19 | -1.04 | -0.48 | 0.54 | 1.75 |
| un4 | 0.74 | 0.46 | -0.77 | -0.22 | 0.81 | 2.01 |
| un5 | 0.76 | 0.41 | -0.82 | -0.26 | 0.77 | 1.97 |
| un6 | 0.91 | 3.11 | 1.88 | 2.44 | 3.46 | 4.66 |
| re1 | 0.94 | 0.90 | -0.33 | 0.22 | 1.25 | 2.45 |
| re2 | 0.72 | 0.54 | -0.69 | -0.13 | 0.90 | 2.10 |
| re3 | 1.07 | 0.81 | -0.42 | 0.14 | 1.16 | 2.36 |
| re4 | 0.80 | 1.76 | 0.53 | 1.08 | 2.11 | 3.31 |
| re5 | 0.67 | 1.42 | 0.19 | 0.75 | 1.77 | 2.97 |
| re6 | 0.93 | 0.57 | -0.66 | -0.10 | 0.92 | 2.12 |
| bt1 | 1.21 | 1.70 | 0.47 | 1.03 | 2.05 | 3.26 |
| bt2 | 1.24 | 0.16 | -1.07 | -0.52 | 0.51 | 1.71 |
| bt3 | 0.67 | 1.32 | 0.09 | 0.65 | 1.67 | 2.87 |
| bt4 | 1.06 | 1.41 | 0.18 | 0.74 | 1.76 | 2.96 |
| bt5 | 0.80 | 1.54 | 0.31 | 0.86 | 1.89 | 3.09 |
| bt6 | 0.96 | 2.26 | 1.03 | 1.58 | 2.61 | 3.81 |

Another indicator of the instrument quality is the person separation reliability and the items separation reliability. Both have the same interpretation as the reliability calculated using Cronbach's alpha: The closer to one, the greater the reliability of the measurement. With regard to the interpretation this means the closer the value to 1, the better the pattern of people's responses, or items endorsements fits the measurement structure (Hibbard, Collins, Baker & Mahoney, 2009). In other words, the separation reliability of people indicates how sure we can be that a person with an estimated ability (in this case perhaps it is better to use affinity to the item) of 2 has, indeed, a greater ability (i.e. affinity to the item) than another person who has an estimated ability (i.e. affinity to the item) of 1. Similarly, the items separation reliability indicates the confidence that an estimated item difficulty 2 has, indeed, a greater difficulty (i.e. lower endorsement level) than another item with estimated difficulty of 1.

Both are calculated using a relationship between the standard error's variance and the mean square error (MSE):

$$\text{Person Separation Reliability} = \frac{\text{var}(\beta\text{'s Standard Error}) - (\text{MSE } \beta)}{\text{var}(\beta\text{'s Standard Error})}$$

$$\text{Items Separation Reliability} = \frac{\text{var}(\delta\text{'s Standard Error}) - (\text{MSE } \delta)}{\text{var}(\delta\text{'s Standard Error})}$$

The separation reliability of the EDTLQ items was 0.99 and the separation reliability of persons was 0.87.

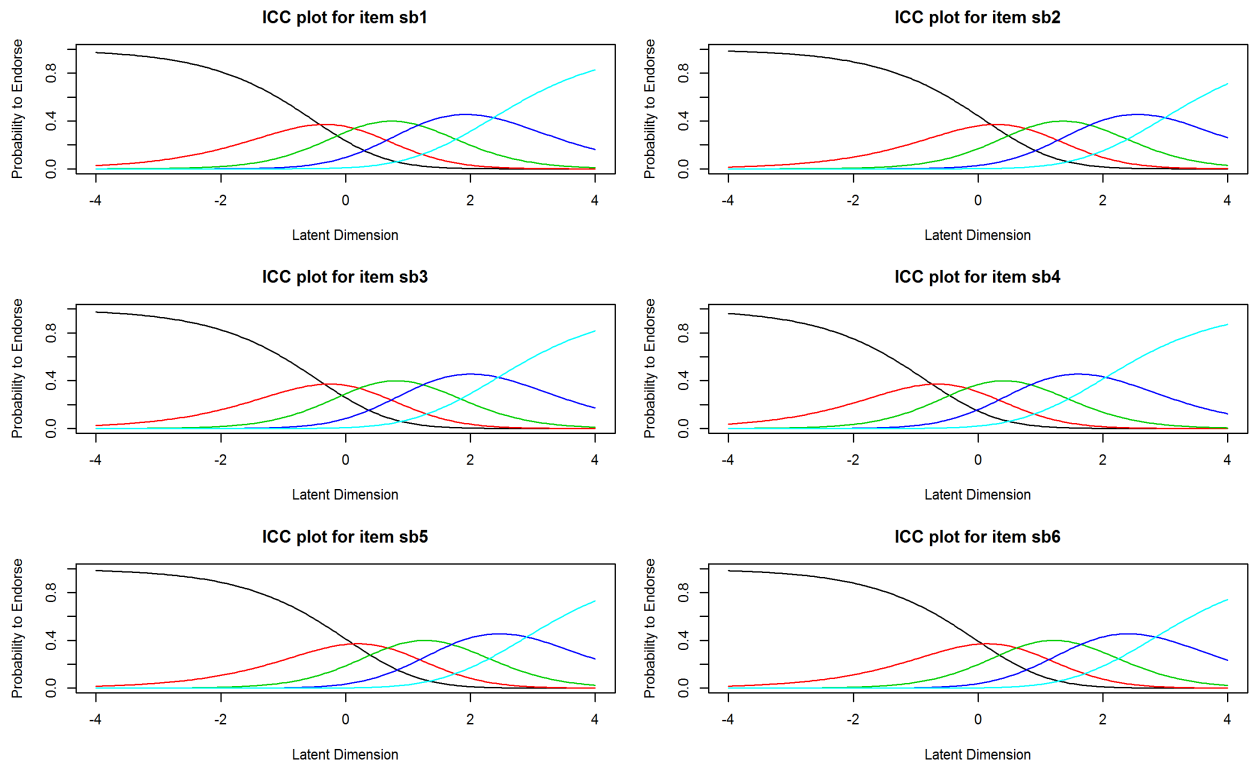


Figure 2. Category characteristic curves for items (statements) from the question “a good study book”. Category one = black line; category two = red line; category three = green line; category four = blue line; category five = light blue line.

Figure 2 shows the category characteristic curves for each item (statement) of the “a good study book” question. The hypothesized ordering of the categories matches the empirical evidence that the first category (value = 1, “least important”) is less difficult to endorse than the second category (value = 2), which is less difficult to endorse than the third category (value = 3), and so on, until the fifth category, the most difficult to endorse (value = 5, “most important”; see also Table 2). Each category, in every item, presents a greater probability of being endorsed than the other categories in at least a small range of the latent variable. Thus, there was no issue related to the ordering of the scales. The same scenario was found in every item of the questionnaire, as can be verified in Table 5. This result means that the response categories of the questionnaire (ranging from “least important” to “most important”) are empirically supported, and do not need to be altered.

Figure 3 shows the parameter distribution of all the items in the EDTLQ. In this figure items ap4 is the easiest to endorse, followed by bt2, un3 and a small cluster of un5, un4, di3, sb4, re2 and re6. Theoretically these items would mostly reflect epistemological level 4 and 5, implying that these levels of epistemological thinking regarding application, good teaching, understanding and discussion correspond closest to what the respondents in the current study feel is important in learning and teaching. At the other end of the scale, the most difficult to endorse statements reflect both the least and the most sophisticated ways of thinking, including the items

- un2, un6 and ap1 reflecting understanding and application focusing on recall and answering exam questions correctly, and
- di2, di6 and bt6 reflecting discussion aimed at hearing different views and the teacher supplying answers and covering the content in the course.

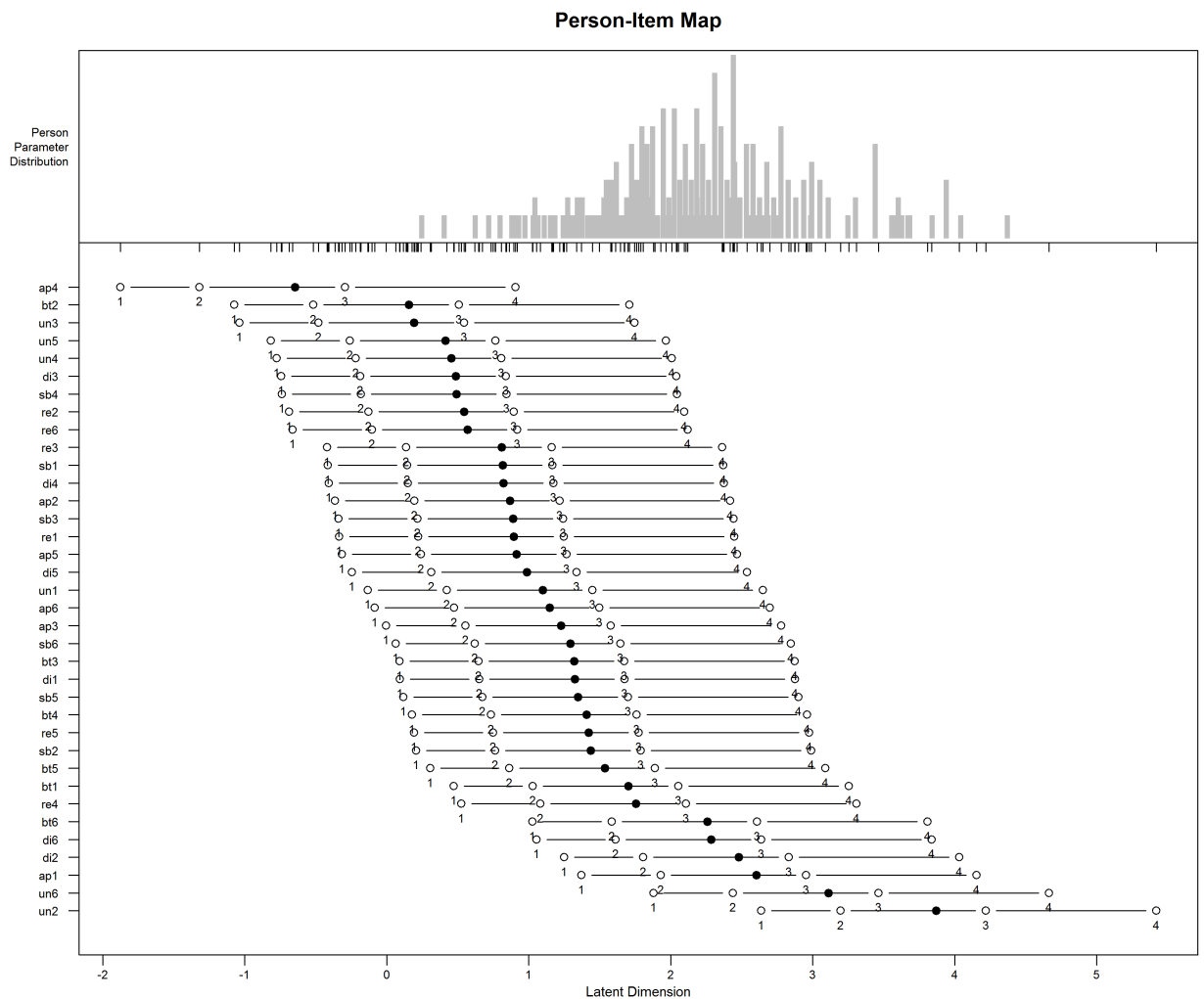


Figure 3. Person-item map, all items of the EDTLQ

This seems to point towards a preference of the respondent group (teachers and staff) as a whole for epistemological thinking at level 4 and to a lesser degree 5, and a rejection of both the most sophisticated level 6, as well as the least sophisticated levels 1 and 2, matching the expected pattern under the consistency hypothesis. To examine in more detail if this pattern is consistent for all the scales, each scale is plotted and discussed separately below.

First we will discuss four of the six scales that seem to behave in a way that confirms the consistency hypothesis where items reflecting the extremes of epistemic thinking are relatively difficult to endorse. These are the scale regarding discussion (di) – discussed in more detail below – and the scales regarding application (ap) and understanding (un) and good teaching (bt). *Figure 4* shows the parameter distribution of the items regarding discussions during a course. Considering only the discussion during course items, item di3 (“different solutions are illustrated from multiple perspectives”) was the least difficult to endorse, while item di2 (“the teacher answers students’ questions”) was the most difficult to endorse. Table 6 links the discussion in course items ordered by the difficulty parameter of the Rasch model to the respective epistemological development level of the items, as per Van Rossum and Hamer’s theory. In this scale the items reflecting the least sophisticated levels of thinking (di1, di6 and di2) are on average the most difficult to endorse, reflecting the pattern found for all the items together.

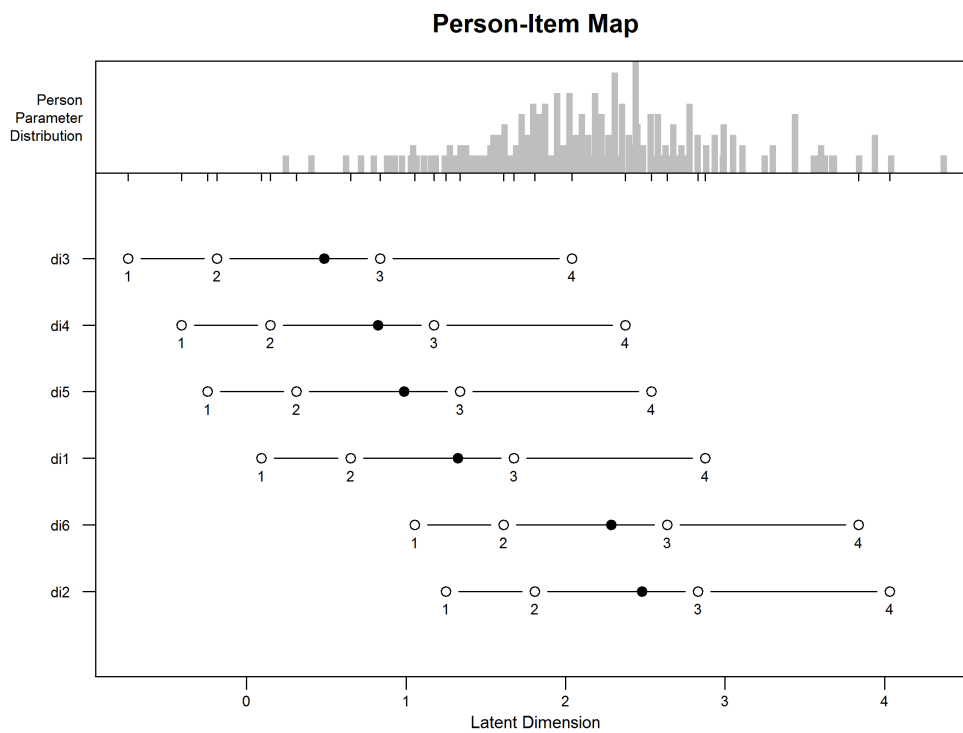


Figure 4. Person-item map, subset of the discussion items (di)

Table 6

Discussion in course items by order of endorsement and epistemological level (Van Rossum & Hamer, 2010)

| Item code | Item parameter | Item endorsement order | Item epistemological Level | Item |
|-----------|----------------|------------------------|----------------------------|---|
| di3 | 0.49 | 1 | 5 | different solutions are illustrated from multiple perspectives. |
| di4 | 0.82 | 2 | 5 | students and teachers learn together and from each other. |
| di5 | 0.99 | 3 | 4 | opinions are based on and/or supported by evidence. |
| di1 | 1.32 | 4 | 2 | the same students don't dominate. |
| di6 | 2.29 | 5 | 3 | you hear all the different views that people have. |
| di2 | 2.48 | 6 | 2 | the teacher answers students' questions. |

A similar pattern can be seen for the items regarding being able to apply knowledge (Figure 5). Items ap4, ap2 and ap5 (reflecting levels 4 and 5) are the least difficult to endorse, while those items reflecting the least sophisticated levels (ap1, ap3; level 2 and 3) and ap6 reflecting the most sophisticated epistemic thinking level were the most difficult to endorse, again reflecting the pattern expected under the consistency hypothesis. The third scale that displays the response pattern expected under the consistency hypothesis is the scale with items regarding the “to understand something means to” issue (Figure 6). Here we see again that the items reflecting higher mid-level epistemic thinking (un3, un5 and un4, reflecting respectively level 4 and level 5 twice, see Table 2) are relatively easy to endorse, whilst those reflecting the less sophisticated



epistemological positions (un1, un6 and un2, reflecting respectively level 3, 2 and 1) are rejected, i.e. difficult to endorse.

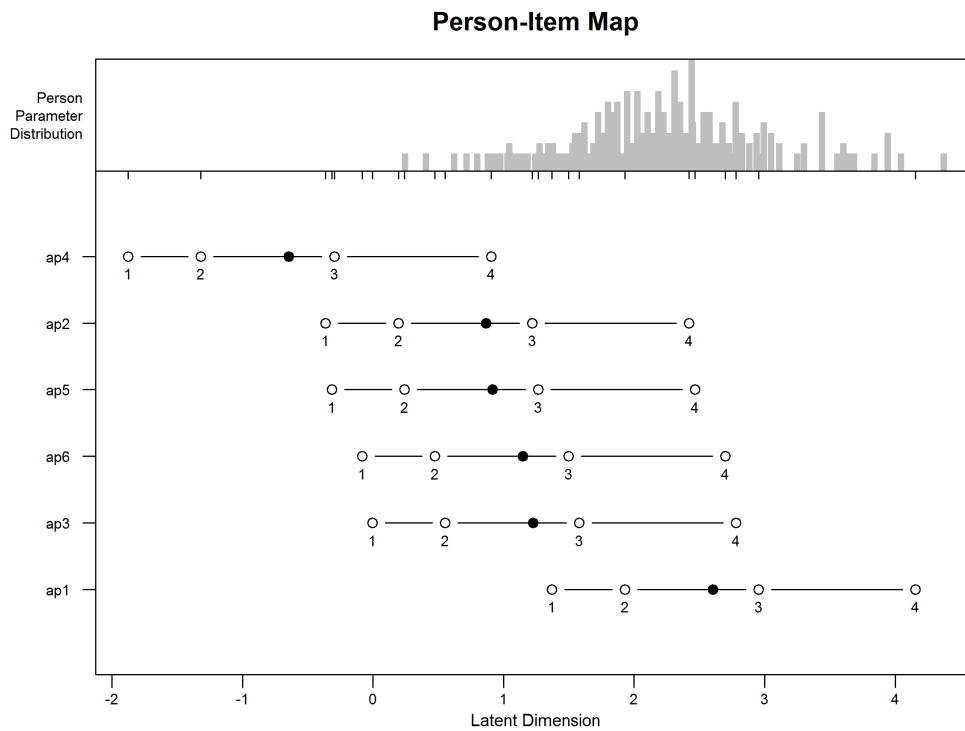


Figure 5. Person-item map, subset of the application items (ap)

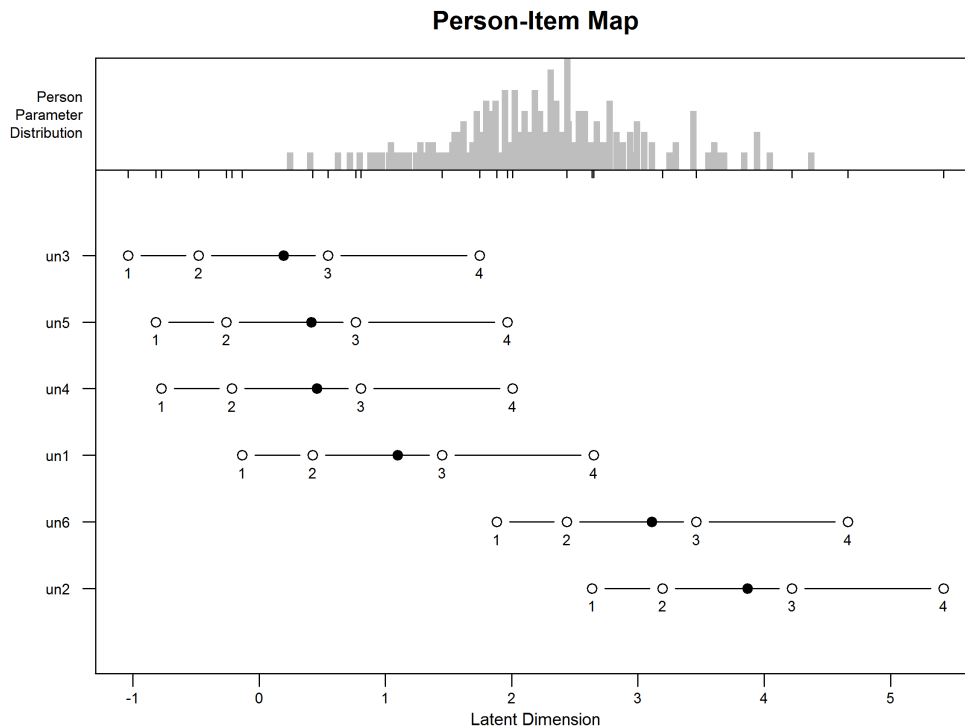


Figure 6. Person-item map, subset of the understanding items (un)

Finally, Figure 7 shows the parameter distribution of the fourth scale that seems to behave as expected – the items regarding the issue “The best teaching ought to ...”. Item bt2 (“inspire the students so



that they are motivated to learn”) was the least difficult to endorse, while item bt6 (“make sure that the whole content of the course is covered”) was the most difficult to endorse. Examining the theoretically linked level of epistemic thinking to the level of endorsement (Table 7), again we see that those items reflecting both the most sophisticated as well as the least sophisticated levels of Van Rossum and Hamer’s model are difficult to endorse, a pattern that is as expected given the rank ordering of the items in combination with the consistency hypothesis.

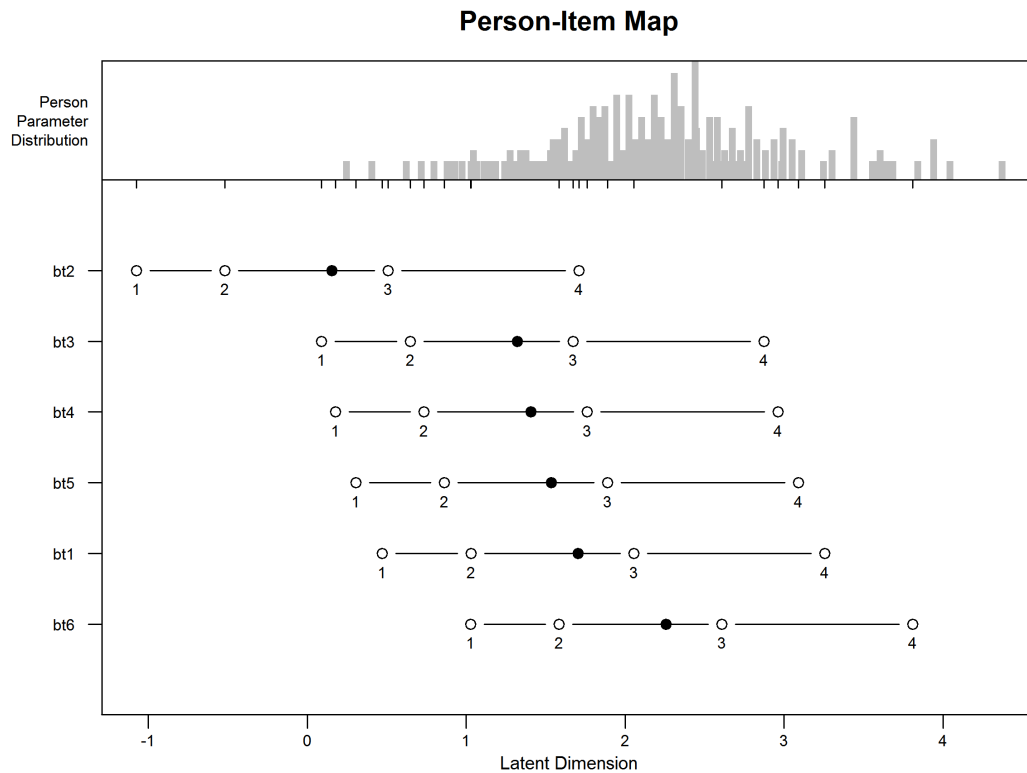


Figure 7. Person-item map, subset of the best teaching items (bt)

Table 7

Best teaching items by order of endorsement and epistemological level (van Rossum & Hamer, 2010)

| Item code | Item parameter | Item endorsement order | Item epistemological level | Item |
|-----------|----------------|------------------------|----------------------------|--|
| bt2 | 0.16 | 1 | 4 | inspire the students so that they are motivated to learn. |
| bt3 | 1.32 | 2 | 3 | focus on application of what is learnt. |
| bt4 | 1.41 | 3 | 5 | make students question their current view on the world. |
| bt5 | 1.54 | 4 | 5 | help students realize the limitations of current knowledge and understandings. |
| bt1 | 1.70 | 5 | 2 | focus on conveying the most important facts and knowledge as clearly as possible |
| bt6 | 2.26 | 6 | 1 | make sure that the whole content of the course is covered. |



The two remaining scales, reflecting the views on a good study book (Figure 8 and Table 8) and responsibility for learning (Figure 9 and Table 9) both do not seem to fit the expectation of the consistency hypothesis.

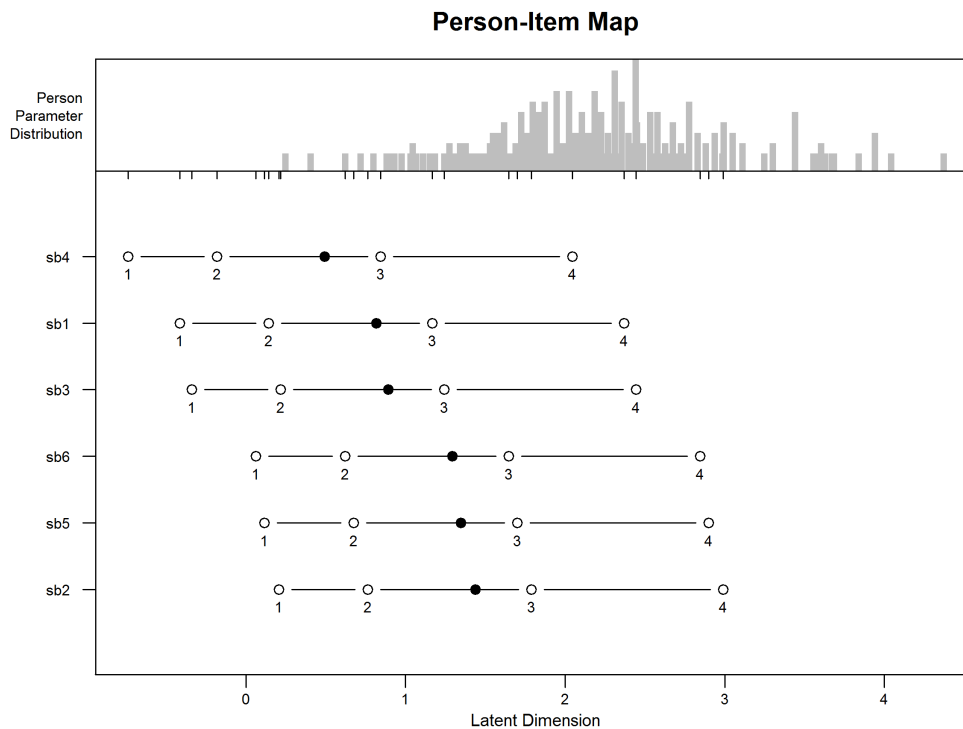


Figure 8. Person-item map, subset of the good textbook items (sb)

Table 8

Good textbook items by order of endorsement and epistemological level (Van Rossum & Hamer, 2013)

| Item code | Item parameter | Item endorsement order | Item epistemological level | Item |
|-----------|----------------|------------------------|----------------------------|---|
| sb4 | 0.49 | 1 | 4 | evokes a critical attitude and invites reflection. |
| sb1 | 0.82 | 2 | 2 | is well structured, gives the main points and the consensus story of the subject. |
| sb3 | 0.89 | 3 | 5 | presents alternative conclusions, shows different interpretations within a subject matter or clarifies different perspectives on the content. |
| sb6 | 1.29 | 4 | 5 | makes students think about the fundamental question of knowledge: "How can we know this?" |
| sb5 | 1.35 | 5 | 4 | provokes and challenges the students' pre-conceptions |
| sb2 | 1.44 | 6 | 3 | provides many up-to-date examples and examples from practice. |



In examining the good study book items, *Figure 8* shows that item sb4 (“evokes a critical attitude and invites reflection”) was the least difficult to endorse, while item sb2 (“provides many up-to-date examples and examples from practice”) was the most difficult to endorse. Examining the level of endorsement by theoretically expected level of epistemic thinking (Table 8) the parameters seem to indicate a clustering inconsistent with the consistency hypothesis. Level 4 thinking represents what the respondent group feels is most important in a good textbook the closest, closely followed by a preference for clear structure (level 2) and clarification of alternative perspectives (level 5). Study books that challenge current thinking and require reflection on the nature of knowledge seem to be more difficult to endorse, although books with many examples from practice (level 3) also seem to be relatively unimportant to the teachers and researchers in this respondent group. The items regarding the nature of a good study book (Van Rossum & Hamer, 2013) are based on student responses only, and have not been similarly confirmed in multiple studies as views reflected in the previous four scales on discussion (di), application (ap), understanding (un) and good teaching (bt).

The final scale is shown in *Figure 9* which gives the parameter distribution of the items regarding the issue “For students to take responsibility for learning means that students...”. Item re2 (“have an active interest in the course and be motivated to learn”) was the least difficult to endorse, while item re4 (“follow their own interests in searching for knowledge”) was the most difficult to endorse. Table 9 shows the responsibility items ordered by the difficulty parameter of the Rasch model and the respective epistemological development level of the items, as per Van Rossum and Hamer’s theory. Item re2 and re6, reflecting “an active interest and motivation to learn” and “fitting what is learned into prior knowledge” are easiest to endorse for the respondent group, whilst items that refer to personal reflection, planning and interests seem to be difficult to endorse and thus seem less important for the teachers and researchers in the respondent group. The response pattern of this scale also does not fit with the expectation under the consistency hypothesis.

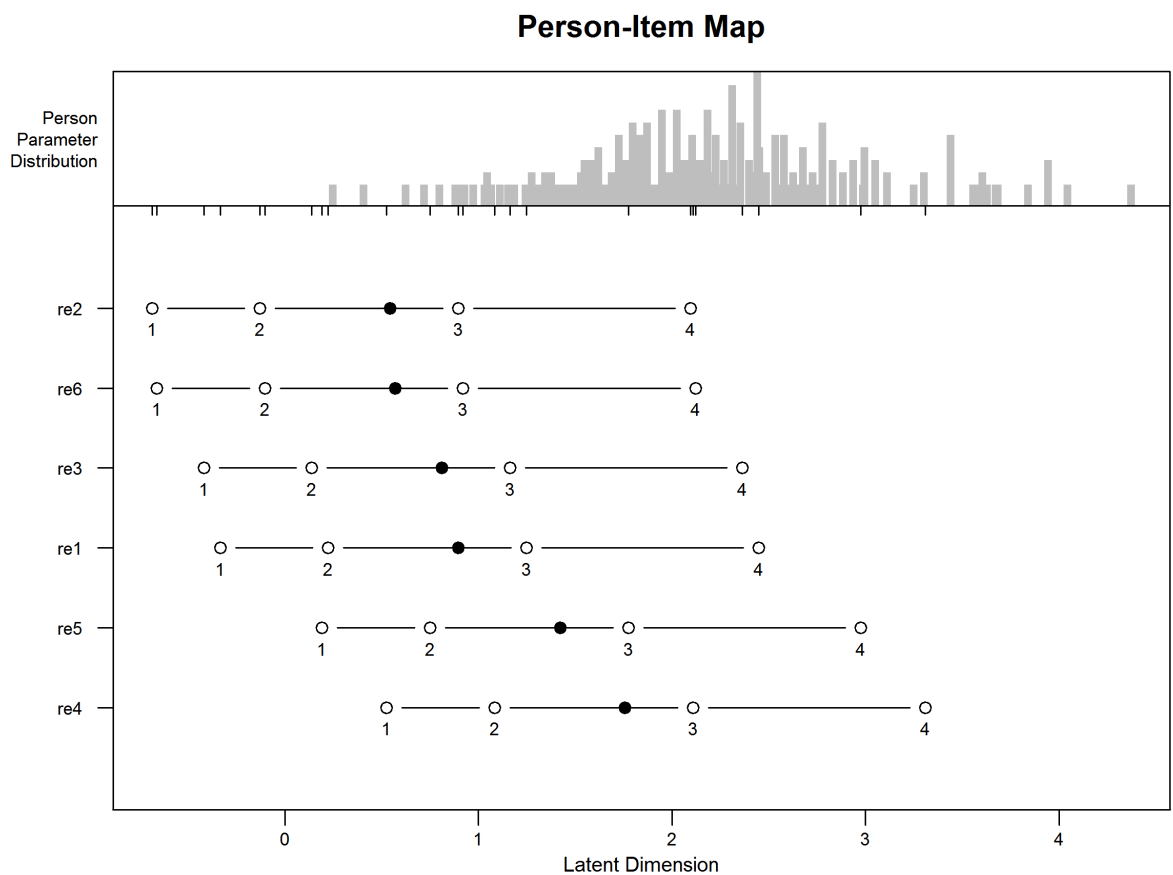


Figure 9. Person-item map, subset of the responsibility for learning items (re)



Table 9

Responsibility items by order of endorsement and assumed epistemological level

| Item code | Item parameter | Item endorsement order | Item Level* | Item |
|-----------|----------------|------------------------|-------------|--|
| re2 | 0.54 | 1 | 3 | have an active interest in the course and be motivated to learn. |
| re6 | 0.57 | 2 | 4 | systematically reflect on what they have learned and how it fits in what they already know. |
| re3 | 0.81 | 3 | 5 | take responsibility for one's own personal development. |
| re1 | 0.90 | 4 | 2 | come to classes well prepared, take notes, review the notes afterwards, and complete assigned tasks. |
| re5 | 1.42 | 5 | 4 | are aware of their strengths, monitoring their progress and making adjustments when necessary. |
| re4 | 1.76 | 6 | 6 | follow their own interests in searching for knowledge. |

*) Please note: the order of these items are not empirically derived but based developmental research literature and work on responsibility issues (Kjellström, 2005; Kjellström & Ross, 2011)

For the overall picture we return to Figure 3 which shows the person-item map including all items, sorted by relative endorsement difficulty, we can construct a similar table for all the items together. The figure shows a level wide range of endorsement with infit measures indicating good fit to the unidimensional Rasch model. The distribution of the levels of the statement is presented in Table 10, and shows that when all statements are taken together, levels 4 and 5 are the easiest to endorse and levels 1 and 2 the most difficult. This response pattern implies that the preferred level of epistemic thinking for this group of teachers is constructivist (level 4 and 5) and teaching by telling and focusing on recall and reproduction is actively rejected.

Table 10

All statements by order of endorsement, based upon Figure 6

| item | Level | Statement |
|------|-------|---|
| AP4 | 4 | be able to solve real life problems by combining knowledge and skills in new ways. |
| BT2 | 4 | inspire the students so that they are motivated to learn. |
| UN3 | 4 | see the connections within a larger context. |
| UN5 | 5 | realize how different perspectives influence what you see and understand. |
| UN4 | 5 | see the underlying assumptions and how these lead to particular conclusions. |
| DI3 | 5 | different solutions are illustrated from multiple perspectives. |
| SB4 | 4 | evokes a critical attitude and invites reflection. |
| RE2 | 3 | have an active interest in the course and be motivated to learn. |
| RE6 | 4 | systematically reflect on what they have learned and how it fits in what they already know. |
| RE3 | 5 | take responsibility for one's own personal development. |
| SB1 | 2 | is well structured, gives the main points and the consensus story of the subject. |



| | | |
|-----|---|---|
| DI4 | 5 | students and teachers learn together and from each other. |
| AP2 | 5 | look at a problem from multiple perspectives. |
| SB3 | 5 | presents alternative conclusions, shows different interpretations within a subject matter or clarifies different perspectives on the content. |
| RE2 | 3 | have an active interest in the course and be motivated to learn. |
| AP5 | 4 | support your own point of view on an issue using evidence and facts. |
| DI5 | 4 | opinions are based on and/or supported by evidence. |
| UN1 | 3 | be able to apply and use it properly. |
| AP6 | 6 | use your knowledge to make the world a better place for those around you (society, human kind and/or nature). |
| AP3 | 3 | be able to solve familiar problems using what you have learnt. |
| SB6 | 5 | makes students think about the fundamental question of knowledge: "How can we know this?" |
| BT3 | 3 | focus on application of what is learnt. |
| DI1 | 2 | the same students don't dominate. |
| SB5 | 4 | provokes and challenges the students' pre-conceptions |
| BT4 | 5 | make students question their current view on the world. |
| RE5 | 4 | are aware of their strengths, monitoring their progress and making adjustments when necessary. |
| SB2 | 3 | provides many up-to-date examples and examples from practice. |
| BT5 | 5 | help students realize the limitations of current knowledge and understandings. |
| BT1 | 2 | focus on conveying the most important facts and knowledge as clearly as possible |
| RE4 | 6 | Follow their own interest in searching for knowledge |
| BT6 | 1 | make sure that the whole content of the course is covered. |
| DI6 | 3 | you hear all the different views that people have. |
| DI2 | 2 | the teacher answers students' questions. |
| AP1 | 2 | pass one's exams. |
| UN6 | 2 | be able to answer test or exam questions correctly |
| UN2 | 1 | you know something by heart |

5. Discussion

Summarizing the results of the analysis of psychometric properties of the six scales comprising the EDTLQ, the confirmatory factor analysis supports the assumption that the six scales together represent a single latent developmental dimension underlying epistemological development that explains the variability in the items, with sufficient infit statistics for all the items. If the assumption was incorrect, the analysis may have resulted in up to six dimensions, one for each issue covered in a scale of the EDTLQ. As this was not the case, the first research question can be answered positively. Examining the factor parameters, with items representing the least sophisticated way of thinking clustering at one end of the dimension and those reflecting more sophisticated epistemological thinking at the other end, the results further seem to imply that



the underlying dimension is somewhat similar to the surface – deep level dimension discussed in earlier works regarding learning strategies (e.g. Entwistle & Ramsden, 1983; Van Rossum & Schenk, 1984).

The rated statements comprising the EDTLQ have satisfactory levels of separation, meaning that the items are characterized by clearly different levels of endorsement and therefore separate the respondents to an acceptable degree, answering the second research question positively. The EDTLQ seems to be a satisfactory tool to measure epistemological views on teaching and learning. The analysis of the category characteristic curves for all of the items in the EDTLQ indicates that there are no response categories that completely overlap, nor are category curves in an unanticipated sequence, indicating that it is not necessary to change the number or sequence of the response categories per item. The design decision to offer a five-point rating scale ranging from “least important” to “important” offers clearly separated response categories. This means that the third research question of this study results in a positive answer as well.

The at first glance puzzling finding from a unidimensional epistemological development perspective, is the fact that the response group of university staff as a whole endorse relatively sophisticated ways of thinking, namely the items reflecting epistemological development levels within the constructivist learning paradigm (levels 4 and 5, see Van Rossum & Hamer, 2010). This result indicates that for this response group, the items reflecting constructivist learning reflect their views on what is important in learning and teaching the best. As discussed in the introduction, a hierarchical inclusive model would predict that items reflecting less sophisticated levels of thinking would be easier to endorse than those reflecting sophisticated ways of thinking, and items reflecting the most sophisticated levels of thinking being endorsed by the fewest of the response group. However, contrary to expectations, in this study it is the items that reflect the least sophisticated levels of thinking that are rejected, i.e. are *not* endorsed, by the larger majority of the response group. Whilst this outcome seems to indicate that the underlying model of epistemological development is incorrect, and so implies that answer to the fourth research question may be negative, there is reason to believe that the response instructions – to indicate how well the statement reflected an important aspect of learning and teaching – led to a different response pattern. As introduced before, Kegan (1994) discusses the consistency hypothesis and states that when circumstances prevent respondents to express their highest level of epistemic beliefs, the pressure felt to “regress” is experienced as unhappiness because “we do not feel ‘like ourselves’” (Kegan, 1994, p. 372). Something similar is discussed by Van Rossum & Hamer (2010) when they describe what they refer to as *Disenchantment* and its counterpart *Nostalgia*. Disenchantment refers to the feeling of disillusionment that students experience when they are exposed to a teaching-learning environment characterized by significantly less epistemological sophistication than their own. This can lead to rebellion or despair (e.g. Yerrick, Pedersen & Arnason, 1998; Lindblom-Ylänne & Lonka, 1999). Nostalgia refers to the wistful hankering to a more traditional learning-teaching environment expressed by students when they are over asked and required to function at an epistemic level too far beyond their understanding (Van Rossum & Hamer, 2010, pp 415 - 426).

Considering the consistency hypothesis, and the request to rate items to reflect the most important level of epistemological development in learning and teaching, it is then in fact not surprising to see that respondents reject items reflecting less sophisticated ways of thinking. Indeed, in Van Rossum and Hamer (2010) many examples are given of students expressing their active rejection of a way of knowing or perception of learning that they feel they have outgrown. Something similar may be taking place in the mind of teachers. However, this does mean that the EDTLQ results of a group reflect the epistemological sophistication level that the majority of the respondent group feel is important to learning and teaching, and when used in a way similar to here does not in fact reflect the epistemological development of a single respondent. To establish what individual respondents feel is important to learning and teaching, reflecting the different preferred levels of epistemological development present in a classroom or lecture hall, will require additional study and analysis, and perhaps different response instructions to those completing the questionnaire.

Examining the six scales of the EDTLQ introduced here, which reflect six different contexts in which an underlying epistemic development may express itself, it seems that four of these five scales behave as can be expected under the consistency hypothesis. These are the scales referring to discussion, application, understanding and good teaching. The scales for views on a good study book and the responsibility for learning scale do not seem to fit the expected pattern.



In the case of the good study book scale, there are two possible explanations that need to be considered and require further study. Firstly, the underlying empirical data for this scale is relatively recent and the analysis of the original data has not been confirmed by further data. Secondly, this scale may provide some insight into the extent to which the teacher responses in this study may be affected by discourse and social desirability. In the pedagogical literature aimed at teachers, significant attention is given to describing constructivist interpretations of application, understanding, the role of discussion in arriving at new knowledge and perspectives and characteristics, providing a rich vein of discourse potentially affecting teacher responses. To the current authors knowledge there is very little of similar discourse available defining characteristics of a good study book. If so, the mixed pattern may reflect the teachers' epistemic thinking more faithfully than the other scales.

Regarding the responsibility scale, again there are at least two reasonable explanations. Firstly, the scale was developed by the use developmental responsibility research from partly another domain (namely responsibility for health) which may illustrate how difficult it is to transfer ideas in one domain to another, and that there may be domain specific patterns (Fischer & Pruyne, 2003). Secondly, and perhaps more importantly, the responsibility items were allocated developmental levels using a model to which they have no empirical link. Hence, the items were interpreted by two of the authors and were allocated a developmental level that may not reflect the actual position on the assumed epistemological development trajectory. Without empirical evidence linking the responsibility items to any of the other scales, it is less clear which level to assign as previous research has shown that a particular statement may be attractive to people different stages of development (Ross, 2008; Loevinger & Hy, 1994).

Remains then the issue that the respondent group of teachers seems to prefer items reflecting relatively sophisticated levels of epistemological development that is usually not observed to this degree and is quite sophisticated even for teachers in higher education (for a review of literature see Van Rossum & Hamer, 2010 chapter 5, 2012). Whilst this seems to indicate that the teachers in this study are characterized by a relatively sophisticated view on learning and teaching, there is another possible explanation: the interpretation of the instructions for rating the statements. Where the example used to illustrate how respondents were to rate the statements referred to a personal view ('I know I have done a good job when...') the items of the five scales were formulated in more general way ('Discussion during a course are good when ...'). This means that we cannot be sure the instruction was interpreted by the teachers as a request to rate each statement in importance to their *personal view on learning*, and therefore that their ratings reflect their personal preference. It is entirely possible that the results reflect '*more* socially constructed discourse and *less* personal ... understanding' (Van Rossum & Hamer, 2010, p. 449). In other words, qualitative responses of teachers are often thick with socially acceptable discourse that at times obscures their personal understanding or preference of the teaching-learning environment (Säljö, 1994, 1997). The pilot of the EDTLQ was performed in Sweden, where the official written curricula focus on skills such as discussing, analyzing and synthesizing rather than on facts and learning by heart. The request to rate statements with regard to their importance to learning and teaching in general may well have activated a preference for items reflecting these written and therefore socially acceptable curricula, above a preference for items reflecting their own beliefs or the reality of the taught curricula.

If this latter interpretation is correct, reformulating the instructions towards an importance to a respondent's own teaching practice may result in a different outcome. Based on theory and earlier experimental findings (Van Rossum & Hamer, 2012, 2013), the findings for the teachers as a group may move slightly towards levels 3 and 4. In addition, the results for students would reflect a significantly less sophisticated preference. As students are less well versed in the academic discourse with regard to teaching, the ambiguous instructions may not have the same effect, which means that the preferred items for the freshmen students in particular would differ considerably with a significant shift towards items reflecting levels 2 and perhaps 3. Although the results of this study are promising, further analyses as well as improvements to the survey, in particular to the rating instruction, and additional data collection are necessary to definitively answer the final research question in this study.



5.1 Methodological reflections

The current study indicates that the unidimensional Rasch model fits the set of scales regarding epistemological preference for a group of university staff. A similar analysis needs to be undertaken for the two groups of students that participated, freshmen and senior (third year undergraduate) students separately, as these two groups may not interpret the scale items in the same way as the teachers and researchers who participated in this study. To establish if the EDTLQ is a viable tool to measure individual epistemological development, further analysis is required to identify specific response patterns for individual participant or participant groups.

To clarify if the results reflect the personal preference in teaching in learning it is necessary to clarify the instructions regarding the rating of the statements. Whilst the categories ranging from least important to most important work well, the instruction above each set of items should include a reference to this personal preference, e.g. "Please rate each statement from 1 (least important) to 5 (most important) to *your own personal views* regarding an ideal learning and teaching environment." Remains the issue of the potential social desirable bias in the teacher responses. When surveying teachers in particular, using inventory statements it will remain difficult to disentangle the use of social acceptable discourse from personal understanding, and as such it is recommended to supplement the EDTLQ with other data collection methods that provide more opportunity for teachers to structure their own response and so provide a more clear window into their personal views.

An interesting finding is that the response group of university teachers seems to be attracted to the more advanced statements, and we probably would not have the same results with high school teachers who may prefer items that reflect levels 2 and 3 (e.g. Brickhouse, 1989, 1990; Martens, 1992; Maor and Taylor, 1995; Hashweh, 1996) which fits the experience in the adult development research field, but hitherto has not been shown statistically. In this sense this result is the most interesting as it enhances the knowledge base within adult development regarding how the statements are endorsed within each item and the whole collection of items. Whilst the main finding, that the items that are easiest to endorse within each scale reflect the midrange levels of epistemological sophistication of the six stage developmental model of learning-teaching conceptions developed by Van Rossum and Hamer (see Table 10) seems contradictory for readers not familiar with adult development theories, both the consistency hypothesis (Kegan, 1994) and the experience in adult development of people endorsing values at their own stage and those slightly more advanced would, is, given the respondent group, exactly what one might expect.

On a more detailed level, there may be issues with individual items or scales. In Table 10 we see that level 4-5 statements are most easily endorsed. Examining the individual items more closely, it is interesting to note in this table is that among the level 5 statements some seem easy to endorse and others more difficult. Three statements that are all about taken perspectives are easily endorsed: "realize how different perspectives influence what you see and understand", "see the underlying assumptions and how these lead to particular conclusions", and "different solutions are illustrated from multiple perspectives". Two level 5 statements that require a more critical awareness and questioning, are more difficult do endorse: "make students questions current view of the world" and "help students realize the limits of current knowledge and understanding". A possible explanation would be that the easier to endorse items can both be endorsed by respondents viewing learning and teaching from a multiplistic perspective (L3) where all opinions are equally valid (Perry, 1970), and by respondents viewing learning and teaching from a relativist perspective (L5), where opinions can differ in validity and in range of supporting evidence, but judging opinions is less acceptable, as in this way of thinking the focus is not on uncovering the most valid opinion, but on empathically understanding the point of view of others (Van Rossum & Hamer, 2010). This means that the easily endorsed items that describe multiple perspective taking need amending, to make them more clearly either level 3 or level 5.

Further, there are some individual items that seem to not fit the model. In particular this is the case for the items regarding the responsibility for learning in particular. As is indicated above, these items were not developed using empirical data from Van Rossum and Hamer (2010). These items seem to reflect the Swedish written curriculum and as such are probably part of the socially acceptable discourse for the respondents in this study. In addition, some items are currently assigned to the lowest level where they apply, e.g. re5 "are aware of their strengths, monitoring their progress and making adjustments when necessary"



reflecting meta-cognition and self-monitoring, whilst these items could well be endorsed by respondents with more sophisticated levels of thinking.

Finally, there could be cultural differences in understanding the items. Van Rossum and Hamer collected their data from Dutch students. Although in their 2010 review, Van Rossum and Hamer explicitly address cultural differences with regard to their model, the understanding of individual items could of course differ. Such cultural differences could include the afore mentioned emphasis in the written curricula on skills such as discussing, analyzing and synthesizing rather than focusing on facts and recall, but perhaps there are other less obvious cultural influences that will only become manifest if the EDTLQ is piloted in other countries and with other respondent groups. An obvious choice given the countries represented by the authors would be an English language pilot, and perhaps a translation into Portuguese. In addition, respondent groups could include teachers and students in secondary education.

Keypoints

- The EDTLQ measures a single latent developmental dimension underlying epistemological development
- The response alternatives showed satisfactory levels of separation which means that the alternatives that are rated do not need to be altered
- The EDTLQ successfully includes items reflecting rare complex epistemic thinking not captured in similar questionnaires before
- As analysed here, the EDTLQ results reflect the dominant epistemic position of the response group. In this study, the response group (HE teachers) seems to prefer items reflecting relatively sophisticated levels of epistemological development that is usually not observed even in higher education, which has implications for construction and instructions of questionnaires.
- The EDTLQ has a potential to be used as an apt tool for teachers to monitor the development of students as well as to offer professional development opportunities to the teachers

Acknowledgments

We would like to thank Kristian Stålné, University of Malmö, Sweden for his valuable contribution to constructing the items and collecting the data. Further, we would like to thank the reviewers of FLR for their insightful comments which helped to make our paper more concise and informative.

References

- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika* 43(4), 561–574.
- Andrich, D. (1988). *Rasch Models for Measurement*. Newbury Park, CA: Sage.
- Andrich, D. (2004). Controversy and the Rasch model: A characteristic of incompatible paradigms? *Medical Care*, 42(1), 7 – 16.
- Andrich, D. (2011). Rating scales and Rasch measurement. *Expert Review Pharmacoeconomics Outcomes Reserach*, 11(5), 571–585. doi: 10.1586/erp.11.59
- Arum, R., & J. Roksa, (2011). *Academically adrift: Limited learning on college campuses*. Chicago, city, 2 letter state: University of Chicago Press.
- Barzilai, S., & Eshet-Alkalai, Y. (2015). The role of epistemic perspectives in comprehension of multiple author viewpoints. *Learning and Instruction*, 36, 86 – 103. doi: 10.1016/j.learninstruc.2014.12.003



- Barzilai, S., & Weinstock, M. (2015). Measuring epistemic thinking within and across topics: A scenario-based approach, *Contemporary Educational Psychology*, 42, 141-158. doi: 10.1016/j.cedpsych.2015.06.006
- Baxter Magolda, M.B. (1992). *Knowing and reasoning in college*. San Francisco, CA: Jossey-Bass.
- Baxter Magolda, M.B. (2001). *Making their own way*. Sterling, VA: Stylus Publishing.
- Belenky, M.F., Clinchy, B.M., Goldberger, N.R., & Tarule, J.M. (1986 ed. 1997). *Women's ways of knowing: The development of self, voice and mind*. New York: Basic Books.
- Bentler, P.M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107(2), 238-246. doi: dx.doi.org/10.1037/0033-2909.107.2.238
- Bentler, P.M., & Bonett, D.G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88, 588-606. doi: 10.1037/0033-2909.88.3.588
- Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ US: Lawrence Erlbaum Associates Publishers. doi: 10.1111/j.1745-3984.2003.tb01103.x.
- Bråten, I. and Strømsø, H.I. (2004). Epistemological beliefs and implicit theories of intelligence as predictors of achievement goals. *Contemporary Educational Psychology*, 371-388. doi: 10.1016/j.cedpsych.2003.10.001
- Brickhouse, N.W. (1989). The teaching of the philosophy of science in secondary classrooms: case studies of teachers personal theories. *International Journal of Science Education*, 11(4), 437-449. doi: 10.1080/0950069890110408.
- Brickhouse, N.W. (1990). Teachers' Beliefs About the Nature of Science and Their Relationship to Classroom Practice. *Journal of Teacher Education*, 41(3), 53-62. doi: 10.1177/002248719004100307.
- Browne, M.W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In: Bollen, K. A., Long, J. S. (Ed.) *Testing structural equation models* (pp.136 -162). Newbury Park: Sage.
- Commons, M.L., & Ross, S. N. (2008a) Special Issue: Postformal Thought and Hierarchical complexity. *World Future: Journal of General Evolution*, 64(5-7): p. 297-562. Doi: 10.1080/02604020802301121.
- Commons, M.L., & Ross, S. N. (2008b), What postformal thought is, and why it matters. *World Future: Journal of General Evolution*, 64(5-7): p. 321-329. doi: 10.1080/02604020802301139.
- Commons, M.L. (1989) *Adult development. Vol. 1, Comparisons and applications of developmental models*. New York: Praeger. xii.
- Commons, M.L. (1990) *Adult development. Vol. 2, Models and methods in the study of adolescent and adult thought*. New York: Praeger.
- Dawson, T.L. (2006), The meaning and measurement of conceptual development in adulthood. In C.H. Hoare *Handbook of adult development and learning*, Oxford University Press: Oxford. p. 433-454. Doi: 10.1558/jate.v6i1.98
- Dawson, T., Goodheart, E., Draney, K., Wilson, M., & Commons, M. (2010). Concrete, abstract, formal, and systematic operations as observed in a 'Piagetian' balance-beam task series. *Journal of Applied Measurement*, 11(1), 11-23.
- Dawson, T. L., Xie, Y., & Wilson, M. (2003). Domain-general and domain-specific developmental assessments: Do they measure the same thing? *Cognitive Development*, 18, 61-78. Doi: 10.1016/S0885-2014(02)00162-4.
- Demick, J., & Andreoletti, C. (2003) *Handbook of Adult Development*. The Springer Series in Adult Development and Aging. 2003, New York: Springer.
- Demetriou, A., & Kyriakides, L. (2006). The functional and developmental organization of cognitive developmental sequences. *British Journal of Educational Psychology*, 76(2), 209-242. doi: 10.1348/000709905X43256
- Embretson, S.E., & Reise, S. P. (2000). *Item response theory for psychologists*. London: Erlbaum.
- Entwistle, N.J., & Ramsden, P. (1983). *Understanding Student Learning*. New York, NY: Nichols.
- Epskamp, S. (2014). semPlot: Path diagrams and visual analysis of various SEM packages' output. R package version 1.0.1. <http://CRAN.R-project.org/package=semPlot>
- Fischer, K.W., & Pruyne, E. (2003) Reflective thinking in adulthood: Emergence, development, and variation. In J. Demick & C. Andreoletti *Handbook of adult development*, Kluwer Academic/Plenum: New York. p. 169-198.
- Glas, C.A. (2007). *Multivariate and Mixture Distribution Rasch Models*. New York: Springer-Verlag.



- Gow, L., & D. Kember, D. (1990). Does higher education promote independent learning? *Higher Education*, 19, p. 307-322. doi: 10.1007/BF00133895.
- Golino, H.F., Gomes, C.M.A., Commons, M.L., & Miller, P. (2014). The construction and validation of a developmental test for stage identification: Two exploratory studies. *Behavioral Development Bulletin*, 19, 37-54. Doi: 10.1037/h0100589
- Hambleton, R. K. (2000). Response to Hays et al and McHorney and Cohen: Emergence of Item Response Modeling in Instrument Development and Data Analysis. *Medical Care*, 38(9), II-60.
- Hamer, R., & Van Rossum, E.J. (2010). De zes talen in het onderwijs: communicatie en miscommunicatie [The six languages in education: communication and miscommunication]. *Onderwijsvernieuwing*, 17 (November).
- Hamer, R., & van Rossum, E.J. (2016). Students' conception of understanding and its assessment. In E. Cano & G. Ion (Eds.), *Innovative Practices for Higher Education Assessment and Measurement*. (pp. 141-162). HERSHEY, PA: IGI Global.
- Hasweh, M.Z. (1996). Effects of Science Teachers' Epistemological Beliefs in Teaching. *Journal of Research in Science Teaching*, 33(1). 47-63.
- Hibbard, J., Collins, P., Mahoney, E., & Baker, L. (2010). The development and testing of a measure assessing clinician beliefs about patient self-management. *Health Expectations: An International Journal of Public Participation in Health Care & Health Policy*, 13(1), 65-72. Doi: 10.1111/j.1369-7625.2009.00571.x.
- Hoare, C.H., (2006). *Handbook of adult development and learning*. Oxford, UK: Oxford University Press. xviii, 579 p.
- Hornik, K. (2014). The R FAQ. Retrieved from <http://CRAN.R-project.org/doc/FAQ/R-FAQ.html>
- Hu, L.T., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1-55. doi: 10.1080/10705519909540118.
- Irving, P.W., & Sayre, E.C. (2013). Upper-level Physics Students' Conceptions of Understanding. Paper presented at 2012 Physics Education Research Conference. In *AIP Conference Proceedings* (vol. 1513, pp. 98-201). Doi:10.1063/1.4789686.
- Kegan, R. (1994). *In over our Heads: The mental demands of modern life*. Cambridge, MA: Harvard University Press.
- King, P.M., & Kitchener, K.S. (1994). *Developing Reflective Judgment*. San Francisco, CA: Jossey-Bass Publishers.
- King, P.M., & Kitchener, K.S. (2004). Reflective judgment: Theory and research on the development of epistemic assumptions through adulthood. *Educational Psychologist*, 39(1), p. 5-18. doi: 10.1207/s15326985ep3901_2.
- Kjellström, S. (2005). Responsibility Health and The Individual [Ansvar, hälsa och människan – en studie av idéer om individens ansvar för sin hälsa]. Linköping Studies in Arts and Science, nr 318. Linköping University.
- Kjellström, S. & S.N. Ross (2011). Older Persons' Reasoning about Responsibility for Health: Variations and Predictions", *International Journal of Aging and Human Development*, 73(2), 99-124.
- Kjellström, S. & Sjölander, P. (2014). The level of development of nursing assistants' value system predicts their views on paternalistic care and personal autonomy. *International Journal of Ageing and Later Life*, 9(1), 35-68. Doi: 10.3384/ijal.1652-8670.14243
- Kohlberg, L. (1984). *Essays on moral development: Vol. 2. The psychology of moral development* San Francisco: Harper & Row.
- Kuhn, D. (1991). *The Skills of Argument*. Cambridge, UK: Cambridge University Press:
- Linacre J. M. (2002). What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions*, 16 (2), 878.
- Lindblom-Ylänne, S., & Lonka, K. (1999). Individual ways of interacting with the learning environment – are they related to study success? *Learning and Instruction*, 9, 1-18. doi: 10.1016/S0959-4752(98)00025-5.
- Linder, C.J. (1992). Is teacher-reflected epistemology a source of conceptual difficulty in physics? *International Journal of Science Education*, 14(1), 111-121. doi: 10.1080/0950069920140110.



- Loevinger, J., & Blasi, A. (1976). *Ego development: Conceptions and theories* (1. ed.). San Francisco: Jossey-Bass.
- Loevinger, J., & Hy, L. X. (1996). *Measuring ego development* (2. ed.). Mahwah, N.J.: L. Erlbaum Associates.Francisco: Jossey-Bass.
- Mair P., & Hatzinger, R. (2007a). Extended Rasch Modeling: The eRm package for the application of IRT models in R. *Journal of Statistical Software*, 20(9), 1–20.
- Mair P., & Hatzinger, R. (2007b). CML based estimation of extended Rasch models with the eRm package in R. *Psychology Science*, 49, 26–43.
- Mair, P., Hatzinger, R., & Maier, M.J. (2014). eRm: Extended Rasch Modeling. R package version 0.15-4. <http://erm.r-forge.r-project.org/>.
- Maor, D. and Taylor, P.C. (1995). Teacher Epistemology and Scientific Inquiry in Computerized Classroom Environments. *Journal of Research in Science Teaching*, 32(8). 839-354. doi: 10.1002/tea.3660320807.
- Martens, M.L. (1992). Inhibitors to Implementing a Problem-Solving Approach to Teaching Elementary Science: Case Study of a Teacher in Change. *School Science and Mathematics*, 92(3). 150-156.
- Mishra, P., & Kereluik, K. (2011). What 21st century learning? A review and a synthesis. In C. D. Maddux, M. J.Koehler, P. Mishra, & C. Owens (Eds.), *Proceedings of Society for Information Technology & Teacher Education International Conference 2011* (pp. 3301–3312). Chesapeake, VA: AACE.
- Newble, D.I., & Clarke, R.M. (1986). The approaches to learning of students in a traditional and an innovative problem-based medical school, *Medical Education*, 20, 267-273. doi: 0.1111/j.1365-2923.1986.tb01365.x.
- Panayides, P., Robinson, C., & Tymms, P. (2010). The assessment revolution that has passed England by: Rasch measurement. *British Educational Research Journal*, 36(4), 611 626. doi: 10.1080/01411920903018182.
- Paulsen, M.B. and Feldman, K.A. (2005). The Conditional and Interaction Effects of Epistemological Beliefs on the Self-regulated Learning of College Students: Motivational Strategies. *Research in Higher Education*, 46(7), 731-768. doi: 10.1007/s11162-004-6224-8.
- Perkins, D. (1993). Teaching for Understanding. *American Educator: The Professional Journal of the American Federation of Teachers*, 17 (3), 28-35. Document available online <http://www.exploratorium.edu/IFI/resources/workshops/teachingforunderstanding.html>, 24 July 2011.
- Perry, W.G. (1970). *Forms of intellectual and ethical development in the college years: A scheme*. New York, NY: Holt, Rinehart & Winston.
- Piaget, J. (1954). *The construction of reality in the child*. New York: Basic Books.
- Pornprasertmanit, S., Miller, P., Schoemann, A., & Rosseel, Y. (2014). semTools: Useful tools for structural equation modeling.. R package version 0.4-6. <http://CRAN.R-project.org/package=semTools>
- Qian, G. and Alvermann, D.E. (2000). Relationship between Epistemological Beliefs and Conceptual Change Learning. *Reading & Writing Quarterly*, 16, 59-74. doi: 10.1080/105735600278060.
- Qian, G. and Pan, J. (2002). A Comparison of Epistemological Beliefs and Learning from Science Text Between American and Chinese High School Students. In B.K. Hofer and P.R. Pintrich (Eds.). *Personal Epistemology*. (pp. 365-385). Mahwah, New Jersey U.S.A.: Lawrence Erlbaum Ass.
- R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Rasch, G. (1960). *Probabilistic Models for some Intelligence and Attainment Tests*. Danish Institute for Educational Research, Copenhagen.
- Raykov, T. (2012). Scale construction and development using structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 472-494). New York: Guilford.
- Richardson, J.T.E. (2012). Erik Jan van Rossum and Rebecca Hamer – The meaning of learning and knowing. Book review. *Higher Education*, 64 (5), pp 735-738. doi: 10.1007/s10734-012-9518-3.
- Ross, S. N. (2008). Postformal (mis)communications. *World Future: Journal of General Evolution*, 64 (5-7), 530-535. doi: 10.1080/02604020802303952.
- Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2), 1-36. URL <http://www.jstatsoft.org/v48/i02/>.
- Säljö, R. (1994). Minding action; Conceiving of the world versus participating in cultural practices. *Nordisk Pedagogik*, 14(2), 71-80.



- Säljö, R. (1997). Talk as Data and Practice – a critical look at phenomenographic inquiry and the appeal to experience. *Higher Education Research & Development*, 16, 173-190. doi: 10.1080/0729436970160205.
- Schommer, M. (1990). Effects of Beliefs About the Nature of Knowledge on Comprehension. *Journal of Educational Psychology*, 82(3), 498-504. doi: 10.1037/0022-0663.82.3.498.
- Schommer, M. (1994). An Emerging Conceptualization of Epistemological Beliefs and Their Role in Learning. In Garner, R and Alexander, P.A. (Eds.). *Beliefs about Text and Instruction with Text* (pp. 25-40). Hillsdale: Lawrence Erlbaum Associates.
- Schommer, M. (1998). The influence of age and education on epistemological beliefs. *British Journal of Educational Psychology*, 68, 551-561. doi: 10.1111/j.2044-8279.1998.tb01311.x.
- Schreiber, J.B. and Shinn, D. (2003). Epistemological Beliefs of Community College Students and their Learning Processes. *Community College Journal of Research and Practice*, 27, 699-709. doi: 10.1080/713838244.
- Sjölander, P., N. Lindström, A. Ericsson, S. Kjellström, 2014, A pattern recognition method for disclosing different levels of value system from questionnaire data, *Behavioral Development Bulletin*, 19(3), 114-127. doi: 10.1037/h0100596.
- Strijbos, J., Engels, N. & Struyven, K. (2015). Criteria and standards of generic competences at bachelor degree level: A review study. *Educational Research Review*, 14, 18-32). doi:10.1016/j.edurev.2015.01.001.
- Van Rossum, E.J., & Hamer, R.N. (2010). *The meaning of learning and knowing*. Rotterdam, Netherlands: Sense Publishers.
- Van Rossum, E.J., & Hamer R., (2012). Analysing higher education teachers' learning-teaching conceptions with a model of student thinking. Paper presented at the AERA Conference, Vancouver, Canada.
- Van Rossum, E.J., & Hamer, R. (2013). The relationship between students' conception of good teaching and their views on a good textbook. Paper presented at the EARLI Conference, Munich, Germany.
- Van Rossum, E.J., & Schenk, S.M. (1984). The relationship between learning conception, study strategy and learning outcome. *British Journal of Educational Psychology* 54, 73-83. doi:10.1111/j.2044-8279.1984.tb00846.x.
- Vermunt, J.D. (1996). Metacognitive, cognitive and affective aspects of learning styles and strategies: A phenomenographic analysis. *Higher Education*, 31, 25-50. doi: 10.1007/BF00129106.
- Voogt, J, Erstad, O., Dede, C., & Mishra P. (2013) Challenges to learning and schooling in the digital networked world of the 21st century. *Journal of Computer Assisted Learning*, 29, 403-413, doi: 10.1111/jcal.12029.
- West, E.J., (2004) Perry's legacy: models of epistemological development. *Journal of adult development*, 112, 61-70. doi: 10.1023/B:JADE.0000024540.12150.69.
- Wright, B. D., Linacre, J. M., Gustafson, J. E., & Martin-Lof, P. (1994). Reasonable mean-square fit values. *Rasch measurement transactions*, 8(3), 370.
- Yerrick, R.K., Pedersen, J.E., & Arnason, J. (1998). "We're just spectators": A case study of science teaching, epistemology, and classroom management, *Science Education*, 82(6), 619-648. doi: 10.1002/(SICI)1098-237X(199811)82:6<619::AID-SCE1>3.0.CO;2-K.