The following is a pre-publication version of

Andrews, P., & Diego Mantecón, J. (2015). Instrument adaptation in cross-cultural studies of students' mathematics-related beliefs: Learning from healthcare research. *Compare: A Journal of Comparative and International Education, 45*.

DOI: 10.1080/03057925.2014.884346

Page numbers will not match those of the published version

Instrument adaptation in cross-cultural studies of students' mathematics-related beliefs: Learning from healthcare research

Abstract

Much comparative education-related beliefs research has exploited questionnaires developed in one culture for use in another. This has been particularly the case in mathematics education, the focus of this paper. In so doing, researchers have tended to assume that translation alone is sufficient to warrant a reliable and valid instrument for cross-cultural research, prompting concerns that a number of necessary equivalences are unlikely to have been addressed. In this paper we consider the nature of these equivalences before examining the literature of a different field, healthcare research, to synthesise an approach to instrument adaptation that is pragmatic but rigorous. Finally, we demonstrate how this pragmatic approach, incorporating extensive cognitive interviews, enabled us to adapt and refine a mathematics-related beliefs questionnaire, developed in Flanders, for use with students, aged 14/15, in England and Spain. Analyses indicate that the instrument so developed is multidimensional, reliable and cross-culturally valid. Some implications are discussed.

Keywords students' mathematics-related beliefs; survey research; instrument adaptation; cross-cultural research; Spain; England

Introduction

In recent years research has shown that students' mathematics-related beliefs play a significant role in determining their engagement with and subsequent learning of the subject (Ma and Kishor 1997; DeBellis and Goldin 2006; Hannula 2006). Moreover, learner beliefs, or affect, is influenced by a variety of factors, including the contexts in which learners operate, the role of teachers, parents and friends, and day-to-day classroom interactions (Op 't Eynde, De Corte and Verschaffel 2002), implying that the wider culture in which a school is located would have its own influence. This conjecture finds support in evidence that despite similarity of content the classroom manifestation of curricula varies considerably from one cultural context to another (Andrews and Sayers 2013; Clarke 2013) as do teachers' beliefs about the nature of mathematics, its teaching and learning (Andrews, 2007; Cai and Wang 2010). In other words, if teachers' mathematics-related beliefs differ cross-culturally, then it would be reasonable to assume that students' beliefs would vary similarly (Schommer-Aikins 2004). However, despite growing interest in cross-cultural examinations of students' mathematics-related beliefs, whether small scale (Berry and Sahlberg 1996; Mason 2003) or large scale (OECD 2004), relatively little attention has been paid to two key issues. The first concerns the extent to which researchers have considered whether mathematics-related belief structures exist in ways that allow for cross-cultural analysis. Indeed, Alexander (2000) argues that all matters educational are so deeply embedded in the cultural contexts in which they occur as to be beyond the simple survey-operationalised structures typically exploited in cross-cultural belief studies. That being said, studies of this criticised form have shown conceptual consistency in cross-cultural belief-related research (Chan and Elliott 2004; Schommer-Aikins 2004) and highlighted both general and culture-specific belief manifestations (Chen and Zimmerman 2007; Whang and Hancock 1994). The second concerns whether those undertaking such studies have considered the extent to which translation, as the means by which an instrument developed in one cultural context is adapted for use in another, is an appropriate strategy. This is a not insignificant problem in mathematics education research. Even well-known studies such as PISA (OECD, 2005), which undertook a cross-cultural examination of student mathematics affect as part of its major focus on mathematics in 2003, effectively exploited an adaptation of translation as the means by which it made operational internationally an instrument developed first in English and then in French. In this paper we attempt to address both questions and, in so doing, contribute to the methodologies and methods of comparative education research.

A first revision of the mathematics-related beliefs questionnaire

Our awareness of these problems emerged when, some years ago (Diego-Mantecon, Andrews and Op ’t Eynde 2007), we undertook a revision of the Mathematics-Related Beliefs Questionnaire (MRBQ) (Op ’t Eynde De Corte and Verschaffel 2006). Developed for use with students in Flanders, the MRBQ was a research-derived multi-dimensional scale shown to be reliable in measuring Flemish 14 years-old students' beliefs about mathematics and its learning. The MRBQ yielded four broad factors concerning students' beliefs about the role of their own teacher, beliefs about the significance of and their own competence in mathematics, beliefs about mathematics as a social activity and, finally, beliefs about mathematics as a domain of excellence, the latter two were found to be less reliable than originally expected.

Our engagement with the MRBQ was prompted by three objectives. The first was to examine the extent to which such belief systems transcend cultural boundaries; we adapted the instrument for use in England and Spain. The second was to explore potential sub-factors, as the spread of items in each of the four broad factors offered strong indications of their existence. The third was to improve the robustness of the latter two factors. Our first attempt at addressing these objectives (Diego-Mantecón, Andrews, and Op ‘t Eynde 2007) was based on the untested assumption that translation alone was sufficient to guarantee a valid tool for cross-cultural research. However, while we succeeded in clarifying sub-factors in the original structures, we became increasingly aware that our assumptions concerning translation had led to a less effective instrument adaptation than we had initially imagined. Moreover, further examination of specialist literature led us to the understanding that mathematics-related belief structures embraced many more dimensions than were addressed in the MRBQ. Thus, we found ourselves in need of a better warranted and more systematic approach to instrument adaptation.

In this paper, therefore, we do not revisit our understanding of the literature on students’ mathematics-related beliefs, as that is, essentially, unrelated to our objectives. Instead we explore the methodological implications related to the ways in which *instrument adaptation* is undertaken - in this paper *instrument adaptation* refers to any act of cross-cultural instrument development or adaptation of an instrument developed in one cultural context for use in another. To achieve these objectives we undertake several tasks. Firstly, we consider what advice the educational research literature has to offer on instrument adaptation. Secondly, we consider how instrument adaptation has been managed by researchers in the field of mathematics education. Thirdly, we review an education-unrelated field, medicine, for further instrument adaptation insights and, fourthly, we show how those insights facilitated the development of a cross-culturally valid multi-dimensional instrument for the measurement of middle grade students’ mathematics-related beliefs.

As indicated above, a recurrent problem in instrument adaptation is a reliance on literal translations that research has shown may result “in phrases that do not make sense, sentences that are poorly constructed, or even the loss of the original meaning of the item” (Van Widenfeld et al. 2005, 140). They cite, by way of example, an expression from a mental health survey instrument, whereby the expression, *your mind going blank*, became, in Dutch, *een gevoel van leegte*, which was back-translated as *an empty feeling* (Van Widenfeld et al. 2005, 136). Such expressions such, despite their transparency of meaning to a native speaker, are colloquial and unlikely to be rendered sensible by any literal translation. Thus, while translation may be procedurally simple, too often it lacks quality control as, typically, no assessment is made directly of the adequacy, or fitness for purpose, of the original language items (Peña 2007), with the consequence that poorly worded or conceptually weak items, even when subjected to successful back-translation, may go undetected (Smith 2004).

Principles for instrument adaptation in educational research

So, what principles exist for ensuring effective instrument adaptation? It would seem evident that an adapted instrument should be as close to the original as possible, even if we are as yet unsure as to what is meant by closeness. It is also increasingly clear that understanding something of the cultures under scrutiny is crucial in effecting instrument adaptation. In this respect, Osborn (2004), drawing extensively on the earlier work of Warwick and Osherson (1973), presents four principles. The first, conceptual equivalence, concerns the question, do the concepts under scrutiny have equivalent meanings in the examined cultures? For example, notions of teaching style may be familiar to English teachers but are largely unknown in other countries (Osborn 2004). The second, measurement equivalence, acknowledges that concepts may differ in their salience from one culture to another or that some topics may evoke higher degrees of sensitivity than elsewhere. For example, religion is a compulsory element of English children's education but French schools are secular by design. The third, linguistic equivalence, reminds us that equivalent words do not always guarantee equivalent meaning, not least because language as translated is not necessarily language as lived. For example, several European languages have only a single verb for the English verbs teach and learn, not only proving linguistically problematic but also highlighting cultural differences in classroom relationships. Lastly, Osborn discusses the need to ensure that samples are as comparable as far as cultural differences will allow, and have they been selected according to methodologically similar procedures?

In relation to instrument translation, Peña (2007) also presents four principles. Firstly, she writes of the importance of linguistic equivalence as translation, even if accurate, may not be sufficient. For example, languages differ in the use of and importance attributed to articles – in English it is frequently unimportant but in Spanish it is. Secondly, she writes of functional equivalence, or an awareness that translation can lead to ambiguity or incongruity of meaning; the aim should be to ensure that "the instrument and elicitation method allow examination of the same construct" (1257). Thirdly, she discusses cultural equivalence, which entails an understanding not only of the salience of the issues under scrutiny to the target audiences but also the ways in which different cultural and linguistic groups interpret the underlying meaning of an item. Lastly, she writes of metric equivalence and the need to consider item or question difficulty.

In sum, a survey adaptation should "keep the content of the questions semantically similar; within the bounds of the target language, keep the question format similar; retain measurement properties, including the range of response options offered; and maintain the same stimulus" (Harkness et al. 2010, p117). Thus, instrument adaptation, when addressed appropriately, should ensure an adapted instrument that is not only "as similar as possible to the source language version" but also "a conceptually equivalent version that measures the same constructs and in which constructs retain the same meaning" (Van Widenfeld et al. 2005, 136).

Instrument adaptation in mathematics-related belief practice

Despite such clear advice, many mathematics-related beliefs studies have tacitly assumed that survey instruments are unproblematically adapted for cross-cultural use. For example, with respect to students, a number of researchers have taken an instrument developed in one culture and translated it for use in others (Pehkonen 1994; Pehkonen and Tompa 1994; Berry and Sahlberg 1996). In all cases, assumptions were made that translation was an unproblematic and valid process. Interestingly, Mason (2003) evaluated Italian students' mathematics-related beliefs by means of author-translated instruments developed in the US and although her translated instrument yielded "a substantial replication" (Mason 2003, 82) of the US results, a scale relating to word problems proved inconsistent. She comments that, "the literal translation from English to Italian of the term "word problem" is not meaningful, so only the term "problem" was used in formulating the questionnaire items for this study". In so doing, even if she had succeeded in constructing a robust scale for use in the Italian context, her instrument would necessarily have examined a different construct from the one underpinning the original scale.

More recently, Hannula and his colleagues seem to have engaged in a game of academic *Chinese whispers*, whereby an instrument developed in one cultural context for use with a particular respondent group was repeatedly, and without comment, adapted for use in different cultural contexts and with different respondent groups. For example, Hannula et al. (2006) exploited, inter alia, self-confidence measures from a scale developed in the US for school children, to examine Finnish trainee teachers' beliefs about mathematics teaching and learning. Later, Hannula and Laakso (2011) used the Hannula et al. (2006) instrument, augmented by two US-derived scales, with Finnish grade four and grade eight students. More recently, Tuohilampi, Hannula et al. (2013) exploited a simplified version of Hannula and Laakso's (2011) instrument to examine the mathematics-related belief structures of Finnish and Chilean third grade students. Interestingly, in both Hannula and Laakso (2011) and Tuolihampi et al. (2013), the instruments were found to be of limited reliability, yet in neither case did the authors acknowledge the possibility of flaws in their instrument adaptation.

Thus, it seems commonplace for mathematics education researchers to report cross-cultural surveys alongside very limited accounts of instruments adaptation. Of course, just because authors offer incomplete accounts does not mean that instrument adaptation is unacknowledged in the conduct of their research. However, a failure to explicate instrument adaptation may, at best, challenge the provenance of a project's outcomes or, at worst, invalidate them.

A healthcare perspective on cross-cultural instrument development

In trying to understand better how we might facilitate valid instrument adaptation, we turned to the literature on cross-cultural healthcare research, which typically examines the extent to which a health-related phenomenon observed in one culture is repeated in another (Costa, Duggan and Bates 2007) or facilitates the healthcare provision of one group of people, usually a minority, within a larger community (Hanna, Hunt and Bhopal 2005). Importantly, the healthcare literature offers insights into how instrument adaptation may be more rigorously addressed. For example, while both Osborn (2004) and Peña (2007) present warrants for four equivalences, Corless, Nicholas and Nokes (2001), in their review of the literature on cross-cultural adaptations of quality of life questionnaires, identified ten. Thus, it would seem that healthcare researchers set higher benchmark standards than those typically found in social science or mathematics education research. Such expectations are exemplified in a number of research methods syntheses (Geisinger 1994; Guillemin, Bombardier and Beaton 1993; Vallerand 1989) that have identified a six stage model of healthcare-related instrument adaptation (table 1), which goes substantially beyond the translation only approach commonly found in much educational research.

Table 1

Preparation	Project teams should review the literature to clarify the study's aims, identify an appropriate instrument and agree a strategy for adapation (Costa et al. 2007).
Forward translation	This typically involves several bilingual persons, although some studies have used just one (Pekmezovic et al. 2007; Lysyk et al. 2002), who tend to be located either in the target country (Costa et al. 2007; Maillefert et al. 2009) or split between target and original (Lau et al. 2002).
Translation review	Initial translations are compared and reconciled. Typically this is group, including translators and project managers, activity (Geisinger 1994; Lysyk et al. 2002), although it may involve a single translator independent of the original translators (Lau et al. 2002). At this stage, significant attention is paid to both linguistic and conceptual equivalence.
Back-translation	As with forward translation, different projects exploit different numbers of personnel. Back translators tend to be native speakers of the language in which the original instrument had been prepared and fluent in the target language.
Back-translation review	Various back-translations are compared. The agreed version is compared with the original questionnaire and discrepancies resolved, which is typically managed by a group comprising bilinguals and project managers (Costa et al. 2007; Gallasch et al. 2007; Lau et al. 2002; Maillefert et al. 2009; Pekmezovic et al. 2007), who should attend particularly to conceptual equivalence.
Pre-test	This conventionally involves piloting or pre-testing of the instrument with small samples of the target groups (Gallasch et al. 2007; Lau et al. 2002; Lysyk et al. 2002; Pekmezovic et al. 2007).

The six stages of healthcare instrument adaption

A second revision of the MRBQ

When undertaking our first adaptation of the MRBQ we were unaware of the role of decentring in instrument adaptation. That is, we did not consciously adopt phrases or expressions that would increase conceptual equivalence at the expense of translations that may be meaningful in one culture but meaningless in another. This was partly due, we know retrospectively, to a widely held perception that the vocabulary and concepts of mathematics and mathematics education enjoy the same meaning and salience in another. This proved not to be the case.

In our second iteration of our adaptation of the MRBQ (Andrews et al. 2011), focused this time on students in England, Slovakia and Spain, attempts at decentring were undertaken alongside attempts to address several elements of the healthcare model. The preparation stage, involving the lead colleague from each country, led to the agreement that the 40 items of the original MRBQ would be augmented by a further 33, drawn from various sources, that we believed would improve the reliability of the original instrument. These sources included, inter alia, scales used by Kloosterman and Stage (1992) and Pintrich and De Groot (1990). The translation stage, also involving lead colleagues, saw the translation of the English version into Slovak and Spanish. A second colleague in each country then back-translated these versions. In so doing, no attempt was made to undertake a translation review, as funds did not allow for two persons to make the initial translations in each country. The back-translated versions were then discussed at full team meetings and changes made to ensure, as far as was practicable in such circumstances, that the conceptual and linguistic intentions of the original instrument were met in the reconciled translations. Following this, small pilots were undertaken before minor changes were made to the final versions for each country. These new versions were given to 220 students from a typical English secondary school, 250 students from a typical Slovak secondary school and 405 from three typical Spanish secondary schools. In all countries the samples comprised equal numbers of students at ages 11/12 and 14/15, chosen as representing transitional points of a child's schooling irrespective of location. Principal components analyses and correlations calculated on factor scores indicated similar factor structures in each country, although there remained inconsistencies and unresolved conceptual issues concerning the use of items like, *We do a lot of group work in this mathematics class*, which appeared to have greater salience in England than either Slovakia or Spain (Andrews et al., 2011). In sum, while the second iteration yielded an improvement on the first, there remained concerns as to the cross-cultural validity of the instrument.

Unfortunately, the extensive procedures of the healthcare model are typically beyond what is possible within the budgets of most social science researchers, prompting the question, what is it reasonable to assume of social science researchers undertaking cross-cultural instrument adaptations? In addressing this question, we return to the pre-test stage of the healthcare model, where we find studies incorporating some form of patient interview (Costa, Duggan and Bates 2007) or cognitive debriefing (Maillefert et al. 2009) to determine the validity and accessibility of adapted instruments. In the following we consider one particular form, the cognitive interview, as a relatively inexpensive approach with the potential to inform instrument adaptation in the social sciences.

Cognitive interviews

Cognitive interviews typically entail the administration of draft questionnaires to a small sample of people with the purpose of determining the extent to which questions generate the data intended by their authors (Miller 2003; Presser et al. 2004). Their purpose is to identify and overcome potential problems, usually with regard to the ways in which particular questions are posed, the meaning informants glean from them and the processes they employ in answering them (Beatty and Willis 2007; Frost et al. 2007). Since the 1990s cognitive interviews have been used frequently in public policy research like censuses (Goerman 2006) and health care research, although its use in the social sciences is rare (Waddington and Bull 2007).

In general, cognitive interviews typically take one of three forms. The first facilitates participants' articulation of their thought processes through the use of think aloud protocols involving little interviewer intervention (Beatty and Willis 2007; Conrad and Blair, 2009). The second invites informants to rephrase an item in their own words (Baars et al 2005), while the third involves a proactive interviewer guiding the interview by means of various forms of probes (Beatty and Willis 2007), of which our reading of the literature indicates there are at least twelve forms according to whether they are meaning-oriented, process-oriented or recall probes (Goerman 2006); retrospective or concurrent probes (Miller 2003; Presser et al. 2004) and based on scripted or non-scripted protocols (Frost et al. 2007).

Cognitive interviews have been particularly effective in facilitating instrument adaptation. For example, Baars et al. (2005), when developing a children's and adolescents' health-related quality of life questionnaire for use in seven European countries, found that cognitive interviews, undertaken in all participating countries, yielded such detailed feedback on the relevance, age-appropriateness and comprehensibility of each item that unexpectedly high levels of cross-national consensus were achieved. Quittner et al. (2000) adapted a French language quality of life measure for use with US sufferers of cystic fibrosis. They exploited different cognitive probes to determine, inter alia, the extent to which conceptual and semantic problems permeated their initial translation and found substantial areas of confusion. For example, the French, *pleine forme*, with its general implication of wellbeing, was translated as *felt very fit*, which informants interpreted as being related to physical fitness. Thus, despite caveats that interviewers should acknowledge that informants may be culturally conditioned to respond to interviews in unexpected or problematic ways (Goerman 2006), evidence indicates that cognitive interviews, undertaken judiciously, have the propensity to support social science researchers in their adaptation of research instruments developed in one cultural context for use in another. In the following, we discuss how such matters played out in our third adaptation of the MRBQ for use in England and Spain.

A third revision of the MRBQ

Our first adaptation of the MRBQ, when we inadvertently neglected the fact the English language items presented in Op 't Eynde, De Corte and Verschaffel (2006) were themselves translations of the original Dutch, involved the second author translating the English of the MRBQ into Spanish. It proved moderately successful but left a number of conceptual and methodological issues hanging. Our second adaptation, which had incorporated elements of decentring, proved more successful but still left some issues unresolved. Significantly, we found ourselves complicit in various forms of cultural insensitivity. For example, we had assumed that all respondents, irrespective of culture, would behave similarly (Rogler 1999).

At this stage, and knowing that resources would not permit us to address all six elements specified in the healthcare model, we made several decisions. Firstly, we would focus attention on adapting our second instrument for use in England and Spain only. Secondly, we would attend more thoroughly to notions of decentring. That is, we would abandon literal translation in favour of one in which conceptual equivalence and the meaning inferred by the reader are privileged. Thirdly, we would exploit cognitive interviews. Fourthly, we would include a wider selection of mathematics education colleagues, both English- and Spanish-speaking, to work on agreeing, translating and refining items through a series of collaborators' meetings. Fifthly, we would, as did the MRBQ, focus on a single age group, that is students in the school year during which they reach age 15.

The collaborators’ meetings highlighted a number of difficulties with respect to instrument adaptation. For example, expressions commonplace in, say, self-efficacy scales like *I am confident that* and *I am certain that* proved problematic with respect to Spanish equivalents, not least because *confident*, *sure* and *certain* imply subtly increasing levels of certainty in English, which our meetings confirmed cannot be accommodated in Spanish. In this respect, the Spanish translation of *I am confident that* is *estar seguro de que*. However, this back-translates as *I am certain that* or *I am sure that*, with the loss of uncertainty implicit in the original. *I am confident of doing something* translates as *confío en hacer algo*. However, *confío* alludes to an implicit sense of hope, which is lacking in the English confident. Furthermore, *I am certain that* translates as *estoy segur*, which back-translates as *I am sure that*. Such matters highlighted well the difficulty of establishing conceptual and linguistic equivalence, particularly with respect to expressions typically employed in self-efficacy scales. Finally, therefore, we adopted the expression, *I am sure that* because it not only translates straightforwardly but also reflects similar levels of certainty in both languages. However, the ambiguity generated by such simple expressions must question the use of many self-efficacy scales in comparative research.

Following this, to identify any unforeseen problems and address any lingering uncertainties with respect to conceptual equivalence, cognitive interviews were undertaken with groups of students in both England and Spain. Nine students were interviewed in each of two Spanish and two English secondary schools located in the regions of Cantabria and Cambridgeshire respectively. The schools, as far as such small numbers allow, were typical state schools with regard to their students' achievement and demographic composition. The students, selected by their teachers to reflect, in equal numbers, the range of mathematics achievement found in their school, included similar numbers of grade 9 boys and girls. All participants were volunteers with their right to withdraw at any time and for no reason made clear to them at the outset. However, while we are confident that the Cantabrian schools reflect the typical Spanish comprehensive school, the same is not the same in Cambridgeshire, where schools in proximity to Cambridge, the home of one of the world's leading research universities, achieve higher than is typical nationally. Thus we acknowledge the possibility that the English data may not be representative of the country as a whole.

Two forms of cognitive interview were undertaken in each school. The first, a group interview, was undertaken in each school. This involved all nine students selected by their respective school and lasted around 30 minutes. This group interview, undertaken independently of the questionnaire, was structured around open questions focused on students' understanding of common mathematical terms and expressions. For example, they were asked, *what do you understand by the term 'maths problem'?* Depending on students' responses, questions, which may have been rephrased, were always followed by requests for students to *give examples of the sorts of maths problems with which you are familiar*. The interviews, which were conducted in the students' schools, entailed students and the second author sitting in a circle. In all four cases students not only talked with assurance and sensitivity to their peers but were able to sustain a discussion requiring little intervention on our part. The second, an individual interview also lasting around 30 minutes, entailed students being invited to respond to the draft questionnaire and to give think-aloud responses to items like, *I am sure I can solve the most complex problems that we are given in class*. Students were asked to explain what they thought the item meant and to articulate their thoughts in relation to its completion – was its meaning clear, if not, why not, and how did he or she think it could it be made clearer? No evidence emerged that students did not understand their task although much of importance with respect to the improvement of the questionnaire was forthcoming.

Results of the interviews

Before presenting the results of the two forms of interview we reflect briefly on our interview processes. Kvale (2006), acknowledging his own earlier failure to address the issue, writes that research interviews embody "a hierarchical relationship with an asymmetrical power distribution of interviewer and interviewee", adding that it "is a one-way dialogue, an instrumental and indirect conversation, where the interviewer upholds a monopoly of interpretation" (Kvale 2006, p. 484). Acknowledging the inevitability of such power imbalances, we believe that our interviews, particularly in the light of their intentions, minimised such one-sidedness. For example, during the group interviews, our roles entailed, as indicated above, asking students about their understanding of common mathematical expressions, followed by prompts to individuals; *what do you think, do you agree with so and so, what do you think your teacher means by such words?* With respect to the individual interviews, our approach was similar. It was essential to the success of our efforts that students were not guided towards a predetermined answer. Indeed, drawing on Ginsburg (1997), we would argue that our conduct of both group and individual interviews encouraged "individual variation" (p. 2), precisely because we allowed control to pass "back and forth between interviewer and child" (p. 39). In short, within a power-constrained context, students' perspectives were unambiguously privileged.

In addition, the language used during the interviews, both by us and the students, was that of the mathematics classroom typical to the student. Indeed, students' sensitivity to the language of mathematics proved immeasurably helpful; help that could not have been obtained differently. That we achieved such results was, we argue, due to our approaching our interviews with no preconceived expectations concerning students' likely mathematical competence. In short, we approached both sets of interviews expecting students to be competent responders, irrespective of gender, ethnicity or prior achievement, as such expectations are known to facilitate both engagement and achievement (Allexsaht-Snider and Hart 2001; Gutierréz 2008).

With respect to our results, it was interesting to note that both sets of interviews highlighted problems with a particular form of wording used in scales generally associated with epistemological beliefs. Statements like *mathematics is useful* were received ambiguously. Some students, both English and Spanish, indicated that it referred to a generic rather than personal sense of utility. For example, David, a Spanish student, talked of mathematics as a tool with which scientists tried to understand the world, while Sarah, an English student commented that "mathematics is used by everyone in their daily lives, like when you go shopping or things like that". Other students interpreted the statement as meaning *mathematics is useful to me* and talked of how they did or could see themselves exploiting mathematics in the future. In almost all cases this was focused on how mathematical competence would facilitate their progression into a world beyond school. Spanish student Luis commented that mathematics is useful because it will help him get to a good university while English student James saw it as something that would get him a better job than someone who couldn't do mathematics. In such examples, the cognitive interviews exposed significant problems in the conventional wording of items well-used in both the literature and our earlier work. In the following, we discuss in detail what the two groups, represented by pseudonyms, had to say.

Spanish interviews

The Spanish students' cognitive interviews indicated, across grades and achievement groups, that the words *problem* and *exercise* were construed consistently and differently from the other. With respect to the former, it emerged that a problem is located in text. A student is expected to read in order to locate the question embedded within it and then devise a solution strategy. In contrast, an exercise was construed as comprising those questions that exploit the minimum words necessary for framing a routine calculation or operation. For example, Alvaro suggested that, "problems have a text, while exercises do not have a text. Also exercises are ready to be done but with problems you have to think and plan them beforehand." Similarly, Sara commented that "… exercises are direct questions and you do not have to think about anything, while you have to think about the problems". More specifically, Beatriz suggested that an exercise included problems like "solve $2x + 3 = 0$ or find x if $3x/2 + 17 = x +2$", while a problem might be "If you have 50 wheels and cars and motorbikes, calculate how many cars and motorbikes you can have".

Occasionally Spanish students appeared to have construed problem and exercise similarly. For example, when confronted with the pilot items *I do mathematics exercises even if I don't like them* and *I do mathematics problems even if I don't like them* a number of students commented that despite differences in wording, they had responded to them similarly. When probed it emerged that these students were not necessarily thinking about the terms problem or exercise but any general tasks given them in class. For example, Antonio, who had earlier distinguished between problem and exercise, illustrated his understanding of the two items by drawing on equations as illustrative of the mathematics he was thinking about when he addressed them, commenting that “I was thinking of equations, because what I don’t like in maths are equations”. His example highlights well how individual construals of terms like exercise or problem can vary according to the context in which they are located. In this case, the items were intended to examine individuals’ persistence in the face of tasks they disliked. Thus, the Spanish cognitive interviews indicated that key expressions like *I don’t like* appear to have a stronger attraction than the words problems or exercises. More generally, this would seem to suggest that readers are drawn more powerfully to words linked to affect than they are to the objects of such affect. Such issues could not have been foreseen as we had taken care to avoid, as suggested in item construction guides (see Fowler and Cosenza 2008), to avoid ambiguity and multiple questions within the single item.

All Spanish students agreed that a mathematics task is work they are instructed to do at home; that is, homework. Also the majority understood the terms exercise and activity to be synonymous. Beatriz, for example, said that “activities and exercises are the same…and maths tasks are homework”, while Elena stated that “maths tasks are the exercises or activities you have to do at home. Activities and exercises are the same to me”. That said, two of the eighteen students construed activities as a broader term that includes exercises; a point made by Maria who commented that “activities are the set of exercises where you revise what you have learnt”.

In summary, the above highlights generally consistent understandings and interpretations of a number of key mathematics-education related terms among the Spanish students. However, the evidence also shows that conceptual consistency is insufficient to guarantee item success. Even though Spanish students were clear in their distinctions between, say, exercise and problem, such distinctions were made problematic by the context of the item, particularly when the context incorporated the affective domain. In such circumstances statements such as *I do mathematics even if I don’t like it* should help avoid such problems.

English interviews

The English data highlighted an inconsistent comprehension of key terms like *exercise* and *problem*. Indeed, roughly half the group offered no distinction between the terms, with the other half clearly thought the terms represented different ideas. Students in the former category indicated that the terms were the same because they both required a solution. Alex, for example, suggested that “both mean the same thing because at the end you have to have a solution or an answer”, while Alison commented that “because problems are to be solved and exercises or questions are to be solved as well”. In similar vein Nick added that “all maths questions are problems that need to be solved, so yes, maths exercises are problems which need to be solved”. Students in the latter group distinguished between the terms. Hannah suggested that problem is a broader category because it includes exercises which require lower cognitive demand. She asserted that problems are “questions involving critical thinking about a statement: or a mathematical puzzle, e.g. Sudoku”, while “exercises relate to operations or calculations, for example solving x-3 = x +2”. In contrast to Hanna, both Tim and Joseph considered problems to be subcategories of exercise. Tim for example suggested that “a maths exercise is a section of work and a problem is a particular part of a question”, while Joseph pointed out that “exercises are made up of many problems, which can often be solved individually”. He cited homework as an example of an exercise and 29700x/7= 1 as an example of a problem.

Other differences were identified. Three of the 18 students indicated that the function of an exercise is to consolidate learning while problems are intended as opportunities to apply knowledge. Sue, for example, commented that “exercises are set to help you understand something and a maths problem tests this knowledge”. Similarly Paul commented that “exercises are just things we carry out to consolidate; problems are made to challenge you”. Other students considered problems to be harder than exercises. Frank commented that “exercises are practices of new methods but problems are harder and take longer”. Such findings not only highlight a lack of consensual understanding among the English students but also make questionnaire design problematic.

Almost all the English students regarded mathematics activities as comprising elements of fun within the more formal tasks they are asked to undertake. For example, Hannah said that “maths activities can involve more thinking out of the box, and can be more interactive and challenging, like a game.” In similar vein, Alex commented that “maths activities are games and generally more enjoyable to do.” Two students indicated additional interpretations. Both Nick and Alex commented that although the term normally indicates a fun element it is also used as a label for exercises or tasks in books as, for example, Nick’s comment regarding “activity 3, page 35”. Interestingly, the English students seemed to construe the expressions mathematics exercise and mathematics tasks as synonymous, with both seemingly related to what their teachers ask them to do. For example, Patrick stated that a “Maths task is something a teacher has told you to do, and…maths tasks and exercises are the same”. Similarly (Alex), stressed that “tasks and exercises are very much the same to me, they are what the teacher sets you do so, like ‘complete page 50’”.

In summary, and acknowledging the lack of consensus among English students, particularly with respect to terms commonly assumed unproblematic in the mathematics education literature, a comparison of the English and Spanish perspectives points to three significant yet probably unforeseen problems with respect to instrument construction. Firstly, the Spanish students, while consistent in their perspectives on key terms, highlighted the extent to which cognitive elements of an item may be subordinated to affective. Secondly, terms that in English convey subtly different meanings find no straightforward translation into Spanish. Thirdly, the extent of the disagreement among the English students suggests that items focused on eliciting student perspectives on frequently used terms like task, exercise, problem or activity may be difficult to construct with any degree of confidence. Thus, instrument development was informed greatly by such matters, to the extent that terms found frequently in the mathematics education literature could not be used.

The final survey instrument

We began this account with reference to the mathematics-related beliefs questionnaire (MRBQ) developed at the University of Leuven, Belgium and, essentially, this is where we finish. The MRBQ was a theory driven instrument yielding four dimensions, two of which were robust and two less so. Over a period of several years, and drawing on a substantial review of the literature, we repeatedly revised the MRBQ to the extent that the final version comprised twelve dimensions, several of which were a consequence of sub-factors emerging from the original instrument, each addressed by 4 items. Importantly, the process outlined above, which we construe as an economically pragmatic response to the healthcare model, enabled us to eliminate the conceptual and linguistic ambiguities and uncertainties embedded in our versions.

Of course, it is insufficient just to assert that the final version was methodologically sound, particularly from the perspective of its cross-cultural validity, and so we close by presenting the results of confirmatory factor analyses (CFA) undertaken on data from students aged 14 and 15 in

Spain and England. In Spain we obtained data from 1270 students from 23 schools near Santander and, in England, 1357 students from 9 schools in proximity to Cambridge. The schools reflected the range of achievement and demographic characteristics of state maintained schools in both areas.

A separate CFA was undertaken on each of the samples in order to examine not only the structural robustness of the underlying theoretical framework, drawn initially from that of the MRBQ, but also its cultural sensitivity. Unlike the exploratory factor analyses we exploited in our earlier iterations, CFA allows both the number of factors and indicator-factor loadings to be specified in advance, thus eliminating the possibility of multiple loadings (Blunch 2008). Moreover, CFA exploits a range of goodness-of-fit tests to determine the adequacy of the proposed structural model of the data; if the model fits the data, then the observed variables measure the proposed constructs. However, caution should be exercised in interpreting the probabilities returned by some of the statistical software as the null hypothesis with CFA is that the model fits the data. In other words, p-values greater than, say, 0.05 are desirable as the intention is not to reject the null hypothesis. Importantly, while each of the goodness-of-fit measures associated with factor analytic models have particular limitations, typically relating to sample size and distributional idiosyncrasy, the totality offer reasonable levels of confidence as to a model's validity (Byrne 2001; Kenny and McCoach 2003).

Table 2

	n	χ^2	df	CFI	TLI	RMSEA
Spain	1170	2426.832	1014	952	948	0.037
England	1300	2087.753	1012	942	934	0.043

Goodness of fit data

The figures reported in table 2 confirm that both data sets, one English and one Spanish, fit the hypothesised model in essentially identical ways. Moreover, the factor loadings shown in table 3, which represent twelve hypothesised constructs each of which drew on four items, highlight several things. Firstly, the very high loadings, particularly when compared with those of the original MRBQ, confirm that each construct is robust and clearly demarcated by its associated items. Secondly, although this is not visible in the table, no item loaded simultaneously on more than one factor, further confirming the robustness of each measure. Thirdly, the Cronbach alphas for the twelve scales ranged from 0.85 to 0.91, indicating highly acceptable levels of reliability. In sum, the analyses confirm that our final adaptation of the MRBQ is a valid and highly reliable multidimensional instrument for cross-cultural measurement of students' mathematics-related beliefs. That being said, it is important for the reader to understand that the labels we have given the constructs are, essentially, arbitrary and reflect our understanding of the literature and, importantly, how the sub-factors of the original MRBQ emerged from our earlier analyses.

Table 3

	Spain	**England**
Teacher directed knowledge		
Maths is about employing the methods in the same way as the teacher	0.757	0.710
Maths is about using the procedures in the same way as the teacher	0.842	0.775
Maths is about following the steps in the same way the teacher shows us	0.798	0.816
Maths is about applying things in the same way as the teacher	0.812	0.838

Strategy use		
Maths is about applying your own strategies	0.755	0.642
Maths is about employing your own tactics	0.818	0.695
Maths is about developing your own plan of action	0.742	0.763
Maths is about planning by yourself which steps to follow	0.667	0.603
Learning through memorising		
I learn maths by doing the same things repeatedly	0.817	0.728
I learn maths by repeating many times what we are taught	0.869	0.811
I learn maths by going over the same things repeatedly	0.826	0.854
I learn maths by practising many times the same things	0.787	0.783
Learning through connecting		
I learn maths by linking the new topics with previous ones	0.809	0.777
I learn maths by connecting the new things to those we have already been taught	0.778	0.786
I learn maths by relating what we are taught to things that we already know	0.703	0.671
I learn maths by identifying things in the new topics that were given in previous topics	0.631	0.544
Usefulness: career improvement		
Maths will help me to find a job	0.819	0.820
Maths will improve my opportunities in the future	0.850	0.839
Maths will help me to earn a living	0.882	0.859
Maths will increase my employment possibilities	0.873	0.810
Usefulness: Real world functioning		
Maths is useful in my daily life	0.788	0.729
Maths is necessary in my everyday life	0.877	0.837
Maths is helpful to me in real life	0.830	0.918
I need maths in different situations of my life	0.765	0.772
Goal in maths education; marks		
My only interest in maths is the results of the exams	0.799	0.816
Grades are the most important thing for me in maths	0.889	0.819
In maths I only care about the marks I get in the exams	0.924	0.900
The only relevant thing for me in maths is grades	0.876	0.875
Self-efficacy		
I am sure I can understand everything we are taught in maths	0.825	0.819
I am sure I can learn everything we are taught in maths	0.855	0.822
I am sure I can comprehend everything we are given in maths lessons	0.875	0.794
I am sure I can take on board everything we are given in maths lessons	0.849	0.815
Self-concept		
Maths is very easy for me	0.804	0.835
I am very good at maths	0.886	0.878
I am very quick at maths	0.845	0.800
I am very skilled at maths	0.879	0.919

Enjoyment of mathematics		
I like maths	0.866	0.851
I enjoy maths	0.902	0.963
It is a pleasure for me to learn maths	0.771	0.790
I love maths	0.901	0.857
Effort		
I put a lot of effort into maths	0.790	0.833
I work very hard in maths	0.850	0.883
I push myself hard in maths	0.909	0.838
I try my best in maths	0.851	0.867
Anxiety		
I become nervous when I do maths	0.859	0.872
When I do maths I become tense	0.867	0.881
I get stressed easily when I do maths	0.760	0.808
Doing maths makes me uneasy	0.721	0.840

Factors and item loadings

Discussion

Developing an instrument for cross cultural research is a complex process, a complexity frequently overlooked in comparative education studies. This study has addressed issues in both methodology and methods in comparative studies of students' mathematics-related affect. From the methods perspective, our pragmatic adaptation of the healthcare model, which offers an economic but rigorous approach to instrument adaptation, yielded a cross-culturally valid and reliable instrument for measuring students' mathematics-related beliefs. In particular, the use of cognitive interviews confirmed how some items invoke substantially different responses from what their authors intended; findings that resonate well with, say, Da Silva et al.'s (2006) measurement, exploiting an adaptation of an instrument developed by the World Bank, of social capital in Peru and Vietnam. Importantly, with respect to issues of wider concern, our cognitive interviews and collaborators' meetings highlighted particular problems linked to the adaptation of instruments developed in English, and especially those employing the vocabulary of uncertainty. That is, due to its wider vocabulary allowing for subtle gradations of meaning, English presents substantial problems for colleagues working in other languages. From the methodological perspective, this study has indicated, at least as far as two diverse European cultures are concerned, a similarity of students' mathematics-related belief structures. In other words, while it would be foolish to argue that all matters educational are not deeply embedded in the cultural contexts in which they occur (Alexander 2000), our evidence confirms the viability of cross-cultural beliefs-related survey instruments. That being said, it would be wise to acknowledge that we adapted an earlier, theory-derived, instrument developed in Flanders. Thus, despite the belief similarity identified above, any instrument can only evaluate those constructs embedded in it and may miss local particularities. In other words, the efficacy of any survey is reliant on the relevance of the constructs it incorporates.

From the perspective of mathematics education, this study has highlighted two key issues of importance for future work in the field of mathematics-related beliefs, whether comparative or not. The first is that many terms common in mathematical discourse remain ambiguous, even within a single cultural context. Words like *problem*, *exercise*, *activity* and *task*, despite a perceived common understanding in the mathematics education literature, have no consistent meaning in the context of the English mathematic classroom. That said, they may have consistent meanings in the Spanish context, which alludes to the second key issue. Just because a term is understood to mean the same thing in one cultural context and is, therefore, distinguishable from another term with a different but related meaning, does not mean that researchers can assume cross-cultural agreement.

In closing, we note that instrument adaptation requires attention to a number of key equivalences which many researchers, including ourselves, have earlier failed to address. It may also mean that researchers may have to compromise in their choice of constructs; our emergent understanding of the problematics of mathematics vocabulary previously assumed unproblematic, particularly that of uncertainty, limited our ability to create a self-efficacy scale. However, despite such problems, it is possible, as we have shown, to construct cross-culturally valid instruments that have the facility to evaluate with accuracy and reliability fundamental constructs of mathematics education.

References

Alexander, R. 2000. *Culture and pedagogy: international comparisons in primary education*. Oxford: Blackwell.

Allexsaht-Snider, M., and L. Hart. 2001. 'Mathematics for all': How do we get there? *Theory Into Practice,* 40, no. 2: 93-101.

Andrews, P. 2007. The curricular importance of mathematics: A comparison of English and Hungarian teachers' espoused beliefs. *Journal of Curriculum Studies,* 39, no. 3: 317-338.

Andrews, P., J. Diego-Mantecón, P. Vankúš, and P. Op 't Eynde. 2011. Construct consistency in the assessment of students' mathematics-related beliefs: A three-way cross-sectional pilot comparative study. *Acta Didactica Universitatis Comenianae - Mathematics* 11: 1-25.

Andrews, P., and J. Sayers. 2013. Comparative studies of mathematics teaching: does the means of analysis determine the outcome? *ZDM*, 45, no. 1: 133-144.

Baars, R., C. Atherton, H. Koopman, M. Bullinger, and M. Power. 2005. The European DISABKIDS project: development of seven condition-specific modules to measure health related quality of life in children and adolescents. *Health and Quality of Life Outcome,* 3, no. 70.

Beatty, P., and G. Willis. 2007. Research synthesis: the practice of cognitive interviewing. *Public Opinion Quarterly* 71, no. 2: 287-311.

Berry, J., and P. Sahlberg. 1996. Investigating pupils' ideas of learning. *Learning and Instruction* 6, no. 1: 19–36.

Blunch, N. 2008. *Introduction to structural equation modeling Using SPSS and AMOS*. Thousand Oaks, CA: Sage.

Byrne, B. 2001. *Structural equation modelling with AMOS: basic concepts, applications, and programming*. Mahwah: Erlbaum.

Cai, J., and T. Wang. T. 2010. Conceptions of effective mathematics teaching within a cultural context: perspectives of teachers from China and the United States. *Journal of Mathematics Teacher Education*, 13 no. 3: 265-287.

Chan, K., and R. Elliott. 2004. Epistemological beliefs across cultures: Critique and analysis of beliefs structure studies. *Educational Psychology,* 24, no. 2: 123-142.

Chen, P., and Zimmerman, B. 2007. A Cross-National Comparison Study on the Accuracy of Self-Efficacy Beliefs of Middle-School Mathematics Students. *The Journal of Experimental Education*, 75, no. 3: 221-244.

Clarke, D. 2013. Contingent conceptions of accomplished practice: the cultural specificity of discourse in and about the mathematics classroom. *ZDM*, 45 no. 1: 21-33.

Conrad, F., and J. Blair. 2009. Sources of error in cognitive interviews. *Public Opinion Quarterly* 73, no. 1: 32-55.

Corless, I., P. Nicholas, and K. Nokes. 2001. Issues in cross-cultural quality-of-life research. *Journal of Nursing Scholarship* 33, no. 1: 15-20.

Costa, F., C. Duggan, and I. Bates. 2007. A systematic approach to cross-cultural adaptation of survey tools. *Pharmacy Practice* 5, no. 3: 115-124.

De Silva, M., T. Harpham, T. Tuan, R. Bartolini, M. Penny, and S. Huttly. 2006. Psychometric and cognitive validation of a social capital measurement tool in Peru and Vietnam. *Social Science & Medicine* 62, no. 4: 941-953.

DeBellis, V., and G. Goldin. 2006. Affect and meta-affect in mathematical problem solving: A representational perspective. *Educational Studies in Mathematics* 63, no. 2: 131–147.

Diego-Mantecón, J., P. Andrews, and P. Op ‘t Eynde. 2007. Refining the mathematics-related beliefs questionnaire (MRBQ). In *Proceedings of the Fifth Congress of the European Society for Research in Mathematics Education* (CERME5), ed. D. Pitta-Pantazi and G. Philippou, 229-238. Larnaka: ERME.

Fowler, F., and C. Cosenza. 2008. Writing effective questions. In *International handbook of survey research*, ed. E. De Leeuw, J. Hox and D. Dillman, 136-160. New York: Lawrence Erlbaum.

Frost, M., B. Reeve, A. Liepa, J. Stauffer, and R. Hays. 2007. What is sufficient evidence for the reliability and validity of patient-reported outcome measures? *Value in Health* 10: S94-S105.

Gallasch, C.,N. Alexandre, and B. Amick. 2007. Cross-cultural adaptation, reliability, and validity of the Work Role Functioning Questionnaire to Brazilian Portuguese. *Journal of Occupational Rehabilitation* 17, no. 4: 701-711.

Geisinger, K. 1994. Cross-cultural normative assessment: Translation and adaptation issues influencing the normative interpretation of assessment instruments. *Psychological Assessment* 6, no. 4: 304-312.

Ginsburg, H. 1997. *Entering the child's mind: The clinical interview in psychological research and practice*. Cambridge: Cambridge University Press.

Goerman, P. 2005. An examination of pretesting methods for multicultural, multilingual surveys: The use of cognitive interviews to test Spanish instruments. In *ZUMA-Nachrichten Spezial Band 12: Conducting Cross-National and Cross-Cultural Survey,* ed. J. Harkness, 67-79. Mannheim: Druck and Kopie Hanel.

Guillemin, F., C. Bombardier, and D. Beaton. 1993. Cross-cultural adaptation of health-related quality of life measures: Literature review and proposed guidelines. *Journal of Clinical Epidemiology* 46, no. 12: 1417–1432.

Gutierréz, R. 2008. A ‘Gap-Gazing’ fetish in mathematics education? Problematizing research on the achievement gap. *Journal for Research in Mathematics Education*, 39 no. 4: 357-364.

Hanna, L., S. Hunt, and R. Bhopal. 2006. Cross-cultural adaptation of a tobacco questionnaire for Punjabi, Cantonese, Urdu and Sylheti speakers: qualitative research for better clinical practice, cessation services and research. *Journal of Epidemiology and Community Health* 60, no. 12: 1034-1039.

Hannula, M. 2006. Motivation in mathematics: Goals reflected in emotions. *Educational Studies in Mathematics* 63, no. 2: 165–178.

Hannula, M., and J. Laakso. 2011. The structure of mathematics related beliefs, attitudes and motivation among Finnish grade 4 and grade 8 students. In *Proceedings of the 35th Conference of the International Group for the Psychology of Mathematics Education*, ed. B. Ubuz, vol. 3: 9-16. Ankara: IGPME.

Harkness, J. A., Villar, A., and Edwards, B. 2010. Translation, adaptation, and design. In *Survey Methods in Multinational, Multiregional, and Multicultural Contexts,* eds. J. Harkness, M. Braun, B. Edwards, T. Johnson, L. Lyberg, P. Mohler, B.-E. Pennell and T. Smith, 115-140. New York: Wiley.

Kenny, D. and D. McCoach. 2003. Effect of the number of variables on measures of fit in structural equation modeling. *Structural Equation Modeling-A Multidisciplinary Journal,* 10, no. 3: 333-351.

Kloosterman, P., and F. Stage. 1992. Measuring beliefs about mathematical problem solving. *School Science and Mathematics* 92: 109-115.

Kvale, S. 2006. Dominance Through Interviews and Dialogues. *Qualitative Inquiry,* 12, no. 3: 480-500.

Lau, A., C. Chang, J. Tai, S. Eremenco, R. Liang, A. Lie, D. Fong, and C. Lau. 2002. Translation and validation of the Functional Assessment of Cancer Therapy-Bone Marrow Transplant (FACT-BMT) Version 4 quality of life instrument into traditional Chinese *Bone Marrow Transplantation* 29, no. 1: 41-49.

Lysyk, M., G. Brown, E. Rodrigues, J. McNally, and K. Loo. 2002. Translation of the Leisure Satisfaction Scale into French: A validation study. *Occupational Therapy International* 9, no. 1: 76-89.

Ma, X., and N. Kishor. 1997. Assessing the relationship between attitude toward mathematics and achievement in mathematics: A meta-analysis. *Journal for Research in Mathematics Education* 28, no. 1: 26-47.

Maillefert, J., M. Kloppenburg, L. Fernandes, L. Punzi, K. Günther, E. Martin Mola, et al. 2009. Multi-language translation and cross-cultural adaptation of the OARSI/OMERACT measure of intermittent and constant osteoarthritis pain (ICOAP). *Osteoarthritis and Cartilage* 17, no. 10: 1293-1296.

Mason, L. 2003. High school students’ beliefs about maths, mathematical problem solving, and their achievement in maths: A cross-sectional study. *Educational Psychology* 23, no. 1: 73-85.

Miller, K. 2003. Conducting cognitive interviews to understand question-response limitations. *American Journal of Health Behavior,* 27, Supplement 3: S264-S272.

Op ’t Eynde, P., E. De Corte, and L. Verschaffel. 2002 Framing students' mathematics-related beliefs: A quest for conceptual clarity and a comprehensive categorization. In *Beliefs: A hidden variable in mathematics education?*, ed. G. Leder, E. Pehkonen and G. Törner, 13–38. Dordrecht: Kluwer.

Op ’t Eynde, P., E. De Corte, and L. Verschaffel. 2006. Epistemic dimensions of students’ mathematics-related belief systems. *International Journal of Educational Research* 45, no. 1: 57–70.

Organisation for economic co-operation and development. 2004. *Learning for tomorrow's world: first results from PISA 2003*. Paris: OECD.

Organisation for Economic Cooperation and Development. 2005. *PISA 2003 technical report*. Paris: OECD.

Osborn, M. 2004. New methodologies for comparative research? Establishing 'constants' and 'contexts' in educational experience. *Oxford Review of Education* 30, no. 2: 265-285.

Pehkonen, E. 1994. On differences in pupils’ conceptions about mathematics teaching. *The Mathematics Educator* 5, no. 1: 3-10.

Pehkonen, E., and K. Tompa. 1994 Pupils’ conceptions about mathematics teaching in Finland and Hungary, *International Journal of Mathematical Education in Science and Technology* 25, no. 2: 229- 238.

Pekmezovic, T., D. Tepavcevic, J. Kostic, and J. Drulovic. 2007. Validation and cross-cultural adaptation of the disease-specific questionnaire MSQOL-54 in Serbian multiple sclerosis patients sample. *Quality of Life Research* 16, no. 8: 1383–1387.

Peña, E. 2007 Lost in translation: Methodological considerations in cross-cultural research, *Child Development* 78, no. 4: 1255-1264.

Pintrich, P., and E. de Groot. 1990. Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology* 82, no. 1: 33-40.

Presser, S., M. Couper, J. Lessler, E. Martin, J. Martin, J. Rothgeb, et al. 2004. Methods for testing and evaluating survey questions. *Public Opinion Quarterly* 68, no. 1: 109-130.

Quittner, A., S. Sweeny, M. Watrous, P. Munzenberger, K. Bearss, A. Nitza, et al. 2000. Translation and linguistic validation of a disease-specific quality of life measure for cystic fibrosis. *Journal of Pediatric Psychology* 25, no. 6: 403-414.

Rogler, L. 1999. Methodological sources of cultural insensitivity in mental health research. *American Psychologist* 54: 424–433.

Schommer-Aikins, M. 2004. Explaining the Epistemological Belief System: Introducing the Embedded Systemic Model and Coordinated Research Approach. *Educational Psychologist,* 39, 1: 19-29.

Smith, T. 2004. Developing and evaluating cross-national survey instruments. In *Methods for testing and evaluating survey questionnaires*, ed. S. Presser, J. Rothgeb, M. Couper, J. Lessler, E. Martin, J. Martin and E. Singer, 431-452. New York: Wiley.

Tuohilampi, L., M. Hannula, V. Giaconi, A. Laine, and L. Näveri. 2013. comparing the structures of 3rd graders’ mathematics-related affect in Chile and Finland. In *Proceedings of the Eighth Congress of the European Society for Research in Mathematics Education (CERME8)*. Antalya: ERME.

Vallerand, R. 1989. Vers une methodologie de validation trans-culturelle de questionnaire psychologique: Implications pour la recherche en langue francaise. *Canadian Psychology* 30, no. 4: 662-680.

van Widenfelt, B., P. Treffers, E. de Beurs, B. Siebelink, and E. Koudijs. 2005. Translation and cross-cultural adaptation of assessment instruments used in psychological research with children and families. *Clinical Child and Family Psychology Review* 8, no. 2: 135-147.

Waddington, P., and R. Bull. 2007. Cognitive interviewing as a research technique. *Social Research Update* 50: http://sru.soc.surrey.ac.uk/SRU50.html (Accessed 20/04/2013)

Warwick, D., and S. Osherson. 1973. *Comparative research methods: an overview*. Englewood Cliffs: Prentice Hall.

Whang, P. and G. Hancock. 1994. Motivation and mathematics achievement: Comparisons between Asian-American and non-Asian students. *Contemporary Educational Psychology,* 19, no. 3: 302-322.