METRIC EQUIVALENCE IN INTERNATIONAL SURVEYS: CULTURAL EDGES

Annica Andersson, Lisa Österling

Stockholm University

This is a critical methodological paper concerning the translation and cultural adaptation processes of an international mathematics education survey questionnaire. Metric equivalence concerns not only language, but also content and activities chosen as indicators in the survey. We here focus the challenges when making cultural, historical and societal considerations when adapting a survey to a new language and cultural context. We conclude that the recommended back translation is not enough to ensure metric equivalence when adapting surveys to a new country. Therefore, we suggest an elaborated method for cultural adaptation. Regarding our survey, this resulted in a survey translation that is better culturally adapted for respondents.

INTRODUCTION

The background for this paper is the now reflected and elaborated answers to an important question posed at the discussion after our presentation at PME37 (Andersson & Österling, 2013): "What were your considerations during the translation process?"

Cross-cultural surveys imply translations of questionnaires to new languages and cultural contexts. To be able to compare results across the borders, the translations need to obtain metric equivalence. The aim of this paper is to document and describe the methodology we developed for translating and adapting a questionnaire from an Australian-Asian context into Swedish language and school culture. We here account for our experiences and critical reflections after the translation and adaptation of the international survey questionnaire within The Third Wave Project, "What I Find Important" (WiFi) (Seah & Wong, 2012), a survey that across cultures investigates what students value as important when learning mathematics. This large-scale quantitative investigation consists of a web-based questionnaire with 89 questions to be distributed to 11 and 15-year old students in 19 different countries. Our task was to translate the questionnaire, into Swedish with possibilities to, first, research what Swedish student value and, second, to be able to make international comparisons.

In a quantitative study, a good measure of values is hard to obtain (see Andersson & Österling, 2013). The problems can be compared to the methodology of attitude surveys, where indicators of attitudes are used instead of posing direct questions (Sapsford, 2007). To obtain metric equivalence, it is crucial that an indicator indicates the same value after a translation. We aim in this study to keep the metric equivalence by conserving the intended meaning of each indicator after translation. Hence, we need to choose either culturally neutral indicators, if such exist, or we need to adapt indicators that conserve the intended meaning across cultural borders.

2014. In Nicol, C., Liljedahl, P., Oesterle, S., & Allan, D. (Eds.) Proceedings of the Joint Meeting of PME 38 and PME-NA 36,Vol. 2, pp. 41-48. Vancouver, Canada: PME.

The WiFi-study is based on value categories from different theoretical frameworks, mainly mathematical values (Bishop, 1988) and cultural values (Hofstede, Hofstede & Minkov, 2010). The value diversity meant a need to differentiate amongst the many dimensions and layers of values that are portrayed in the classroom. To give some examples; Seah & Wong (2012) take the stance that "values are regarded in [the Third Wave project] from a sociocultural perspective rather than as affective factors." This sociocultural perspective may imply that values can be found in relationships, languages and available discourses. Hofstede et al. (2010) instead define values as "the core of culture", explaining that culture reproduces itself and its values through cultural practices. Those practices can be what parents say and do when fostering their children, or what activities teachers choose to do in the classroom. How activities are values is decided by the members of the cultural group.

The questions in the WiFi-survey questionnaire consist mainly of activities from mathematics classrooms. Respondents are asked to answer how important each activity is when learning mathematics. The different activities were chosen as value indicators in the WiFi-questionnaire. Therefore, we need to address cultural practices in the mathematics classroom to validate that the intended meaning of our indicators was culturally stable. In this validation process, we used several methods: repeated pilot tests, interviews with targeted students and educational and historical research to understand the cultural background of Swedish mathematics education.

Historical, Societal and Cultural Background of Swedish Mathematics Education

From the results of WiFi-study we will learn more about what students express as important learning activities in mathematics. To obtain a cultural adaptation while maintaining metric equivalence during translation, we needed deeper knowledge about societal and historical facts that form mathematics educational practices. Otherwise, it is hard to determine what value a value indicator indicates. To give an example, Lundin's (2008) work shows that when Swedish schools became public and mandatory in 1842, teachers had to deal with a large number of children that were the first generation attending school. The first early schoolbooks had two aims; to support the learning of mathematics and support teachers to cope with disciplinary problems. "This need led to the promotion of schoolbooks filled with a large number of relatively simple mathematical problems, arranged in such a way that they (ideally) could keep any student, regardless of ability, busy – and thus quiet – for any time span necessary." (Lundin, 2008, p.376). Mathematics was used as a medium for fostering children. The School Inspectorate's research report (2009) concludes that teachers are still relying on textbooks when planning their teaching, hence trust the textbook to fulfil curriculum objectives. Lundin's (2008) explanation of the historical development might explain the School Inspectorate's (2009) results. This particular way of organising mathematics education is believed to support teachers in managing non-homogeneous student groups so that each student can work according to his/her previous learning and needs. It is likely that parents and students expect mathematics classes to be conducted this way. Hence, working quietly in the textbook has become part of the culture of Swedish mathematics classrooms.

METHODOLOGY

As commonly practiced, the WiFi-study Research Guidelines (not published) suggested translation and back translation as a way of obtaining metric equivalence. However, after having done successful back translations, we conducted a pilot test of the translated questionnaire with a sample of 11-year-old students. It turned out that there was several questions the targeted students did not understand. Therefore, we needed to consider how we best could adapt the questionnaire to a Swedish context, and how best choose contents and activities as indicators that Swedish students are familiar with. A back translation did not serve our purposes. We needed other methods for the cultural and linguistic adaptation.

Exploring methods of cultural adaptation

Translation and back translation can be conducted to investigate problems in the target text. However, this produces limited information of the quality of the target text – which also, as described, became our experience. Harkness, Villard & Edwards (2010) criticizes the use of back translation as a standard method, drawing on research that shows that appraisal of the target text directly is more efficient.

We explored, evaluated and adapted the guidelines for cross-cultural research, published by the Survey Research Center (2010). Harkness et al. (2010) suggests "The TRAPD Team Translation Model" as current best practice. The steps in this model are; **T**ranslation, where two translators make two independent translations; **R**eview, where the translations are compared and refined; **A**djudication, where the translation is separated from review with focus on, amongst other things, a cultural adaptation; **P**ilot test and finally **D**ocumentation of every step in this process. A team should include translator, reviewer and adjudicator. Adjudication is suggested to follow these steps; linguistic mistakes in the translation process, cultural adaptation problems, questions that do not work in the intended group and generic problems from the source version. Each survey is unique, and we adapted this model to suit the circumstances of our project. The frames of this project did not allow for hired professional translator or to organize extensive pilot tests. But we had a team, consisting of three mathematics teachers' educators and researchers. We used the different stages iteratively, and went back to new translations, reviews and adjudications. During this process, we added scoping interviews with students as well as knowledge from earlier educational research to improve the cultural adaptation. Below we describe how this adapted model was used to improve the quality of the translated questionnaire and to keep the metric equivalence.

RESULTS

Results from the adapted TRAPD-process

Scoping interviews: We needed to learn more about how the intended group of students themselves expressed their valuing and interpreted our questions. Semi-structured scoping interviews (Bryman, 2012) were hence conducted. In the translation process, this was intended to help us use students wording and examples in our translation and to facilitate the understanding of the questions.

1st translation: In this stage, the translators, three persons in our case, made a close translation of the WiFi-questionnaire from English to Swedish.

1st review: The translators compared and reviewed each other's translations in review meetings to decide on the best translation. We focused at this stage to keep the translation as close to the original version as possible for a successful back translation.

Back translation: Two persons, who had not previously seen the questionnaire, conducted the back translation from the Swedish translated questionnaire to English.

1st adjudication: In our project, also the adjudication was a team work. We compared the original and the back translated questionnaires and used colour codes to grade the similarities/differences between them. Since the 1st translation was close to the source questionnaire, the back translation was acceptably similar to the source questionnaire.

1st pilot test: In this pilot test, a group of 28 eleven-year-old students were asked to answer the questionnaire, and when doing so, indicating what questions they found difficult to understand or interpret.

2nd adjudication: When analyzing the pilot test, there were too many questions students found difficult to understand. We concluded that we needed to improve the cultural adaptation as well as the adaptation to the intended group. We looked up items in research texts and in the curriculum to check for meaningful and proposed activities in a Swedish context. An example can illustrate the process so far:

Example 1: Q9 focuses ”Mathematics debates”. In the 1st translation, this was easily translated to "*Debatter med matematik*", and the back translation was close enough, "debating maths". However, when trying out the questionnaire in the pilot test, eleven students out of 28 did not understand the question. And when discussing "Mathematics debates" in the 2nd adjudication, not even we as adjudicators were sure about how such a debate is enacted in the classroom. "Mathematics debates" are in the WiFi Research Guidelines (not published) classified as an indicator of valuing openness and exploration. Mathematics debates is not an activity that is common in Swedish classrooms, so out of what it is supposed to indicate, we tried to adapt the indicator, and describe an activity that children could recognize. In the 2nd translation, the question was formulated "*Debattera och ifrågasätta lösningar i matematik*" (Debate and question mathematical solutions), a cultural adaptation so respondents can visualize a situation while still relating to valuing openness and exploration.

Documentation was kept during the whole process of all the different versions of each question. It supported our evaluation of the improvement of quality.

This process made us realize that translation and back translation is not a good instrument to ensure metric equivalence when researching students valuing when learning mathematics. We need to use other methods and decided to take the adaptation one step further. Consequently we followed up the pilot test with interviews of participating students in order to better understand the intended meaning of their answers to some of the questions.

Understanding respondents' intended meaning

Respondents obviously need to understand survey questions. Therefore, we asked them how they interpreted the questions and what their intended meaning was when answering our questions.

Example 2: According to the pilot test a large proportion of students valued Q36 "Practicing with lots of questions" as important or absolutely important. However, Sara, 11, did not. We discuss this result in particular, since it aligns with research results, which show that this is an important trait of Swedish mathematics education.

This question was not hard to understand or to translate. Still, we got contradictory answers in the interviews. We wanted to find out what students valued when they responded that "practising a lot" (*öva genom att göra många uppgifter*) is important or not. Sara, 11, expressed:

Interviewer: - Do you think you need to practise a lot to learn mathematics?

Sara: - Well, if you are already good at it… no!

Her reasoning and intended meaning of this response was more elaborated and very different from what we predicted. She here stated that "good" students don't need to practice that much. However, later in the interview, she gives us examples of mathematical content one always needs to practise a lot, which is practicing the times-tables. She also recognises that there is a different learning process in learning times-tables from learning problem solving, but she cannot express what she finds important for learning problem solving. Her rating of "Practicing with lots of questions" was "neither important nor unimportant". Therefore, using "Practicing with lots of questions" as an indicator becomes hazardous, since respondents make connections and reflections we cannot predict. Interviews with students allowed us to discover some of those unpredicted responses, thus allowing us to problematize conclusions from the data.

3rd adjudication: We worked further on finding expressions and concepts from Swedish classroom contexts. We used previous educational and historical research, as well as our years of experiences as teachers and teacher educators to find the best expressions that could fit classroom cultures and the selected age group of the respondents. At this stage, the team used all information we had gathered to reconsider our translation and adaptation. We used results from the pilot test, from interviews, from a curriculum analysis and the back translation. This method allowed us to evaluate our translation from several perspectives.

2nd translation: We moved away from our initial intention of keeping the target questionnaire (the translated version) as a close translation to pass a back translation. Instead, we put a lot of effort in analyzing what activities that could be the best indicators of the requested value. The use of indicators in the WIFI-study has previously been discussed by Andersson & Österling (2013). We give an example to show how we worked through the whole process.

Example 2: Q11 focuses "Appreciating the beauty of maths" and Q60 "Mystery of maths" were not comprehensible for the Swedish students due to the pilot test. The version we tried out was a close translation. In the 2nd translation we chose to give examples to illustrate what "beauty and mystery of maths" can be. Q11: "*Uppleva att matematik kan vara vacker (som mönster i konst, arkitektur och natur*)" (Experience that mathematics can be beautiful (like patterns in art, architecture and nature) and Q60: " *Undersöka gåtfulla matematikexempel (till exempel kan du lätt mäta en tredjedel av 9 cm exakt med linjal, men en tredjedel av 10 cm går inte att mäta exakt*)", (Exploring enigmatic mathematical examples (e. g. you can measure a third of 9 cm exactly with your ruler, but you cannot measure a third of 10 cm exactly). If those questions were to be back translated, a comparison would say that they are quite different. But the intended meaning is easier for respondents to understand. Therefore, this way of adapting questions to what is familiar of respondents conserves the intended meaning, and thus improves the metric equivalence, since the new question works as an indicator of the values intended.

To sum up, there were a large proportion of questions where the mathematical content and/or the mathematical activities in classrooms were not familiar to Swedish eleven-year-old students. There were also questions that could be interpreted differently, due to cultural differences or due to individual experiences amongst respondents. Therefore, we made some clarifying examples, or even chose a different activity, to try to improve the metric equivalence and construct validity.

CONCLUDING DISCUSSION

Quantitative cross cultural surveys and assessments like TIMSS or PISA are increasingly important aspects of policy making decisions about mathematics education. Those investigations pose the same questions in all countries, since the aim is to compare knowledge between countries.

Recognizing that there are historical and cultural differences between participating countries make it problematic to compare the assessed knowledge, since it is based on the assumption that mathematical content is valued equally everywhere. The WiFi-study is different; it surveys what students find important and does not assess students' mathematical knowledge. But the survey still suffers from the same difficulties, that we are not sure if mathematics or mathematical activities are valued equally across the participating countries. Translating a questionnaire with questions about learning mathematics does not imply only linguistic aspects. The indicators need to be evaluated out of what mathematical content students recognize when being part of the subject of mathematics and what mathematical activities students from different countries or cultures are familiar with.

The WiFi Research Guidelines (not published) suggested translation and back translation. However we could conclude that a successful back translation is not enough to ensure metric equivalence. Having our minds set on how to translate questions so that they would suite the back translation resulted in a too close translation, and respondents in the pilot test did not understand all the questions. Therefore, a back translation did not help us neither with the meaningfulness of item content to each culture, or with the metric equivalence. Instead, an adapted TRAPD-model (Survey Research Centre, 2010) gave us useful tools to improve the cultural adaptation. However, a cultural adaptation cannot be drawn too far without affecting the instrument validity across languages. We had to pay careful attention to maintain the metric equivalence in order to have the possibility of making cross-cultural comparisons of students' values, as intended in the WiFi-project (Seah, 2013). From the results from a finished WiFi-study we can learn more about differences between cultures and values in mathematics learning. However, our dilemma is that at the same time, we depend on some of this knowledge when adapting a proper questionnaire.

Until our larger research study shows us where edges of cultural values can be found in mathematics education, we recommend the other seventeen teams within the WiFi-project, or similar cross cultural projects, to reflect on the translations and cultural adaptations and maybe adopt and adapt further the team translation process. Within the adjudication stages, there are rich opportunities to critically reflect on cultural adaptations through interviews, pilot tests and previous research to improve metric equivalence in cross-cultural research.

References

Andersson, A., & Österling, L. (2013). Measuring immeasurable values. In A. M. Lindmeier & A. Heinze (Eds.), *Proc. 37th Conf. of the Int. Group for the Psychology of Mathematics Education* (Vol. 2., pp. 17-24). Kiel, Germany: PME.

Bishop, A. (1988). *Mathematical enculturation: A cultural perspective on mathematics education*. Dordrecht: Kluwer Academic Publishers.

Bryman, A. (2012). *Social research methods* (4th ed.). Oxford: Oxford University Press.

Harkness, J. A., Villar, A., & Edwards, B. (2010). Translation, adaptation and design. In J. Harkness, M. Braun, B. Edwards, T. P. Johnsson, L. Lyberg, P. P. Mohler, … T. W. Smith (Eds.), *Survey methods in multinational and multiregional contexts* (pp. 117-139). Hoboken, NJ: John Wiley & Sons.

Hofstede, G., Hofstede, G. J., & Minkov, M. (2010). *Cultures and organizations, software of the mind: Intercultural cooperation and its importance for survival* (3rd ed.). New York: McGraw Hill.

Lundin, S. (2008). *Skolans matematik: en kritisk analys av den svenska skolmatematikens förhistoria, uppkomst och utveckling* (Unpublished doctoral dissertation). Acta Universitatis Upsalensis, Uppsala.

Sapsford, R. (2007). *Survey research*. London: Sage Publications.

Seah, W. T. (2013). Assessing values in mathematics education. In A. M. Lindmeier & A. Heinze (Eds.), *Proc. 37th Conf. of the Int. Group for the Psychology of Mathematics Education* (Vol. 3., pp. 193-201). Kiel, Germany: PME.

Seah, W. T., & Wong, N. Y. (2012). What students outside Asia value in effective mathematics learning: a 'Third Wave Project'-research study. *ZDM*, *44*, 1-2.

Skolinspektionen [School Inspectorate]. (2009). *Skolinspektionens Kvalitetsgranskning, Undervisningen i matematik - utbildningens innehåll och ändamålsenlighet: Rapport 2009:5* [Teaching of mathematics - the content and efficiency of education]. Stockholm: Frizes.

Skolverket [The Swedish National Agency for Education]. (2011). *Läroplan för grundskolan, förskoleklass och fritidshem* [Curriculum for the compulsory school, preschool class and the recreation center]. Stockholm: Fritzes

Survey Research Center. (2010). *Guidelines for best practice in cross-cultural surveys.* Ann Arbor, MI: Institute for Social Research, University of Michigan. Retrieved from http://www.ccsg.isr.umich.edu/