REVISTA DE ESTUDIOS E INVESTIGACIÓN
EN PSICOLOGÍA Y EDUCACIÓN
ISSN: 1138-1663; eISSN: 2386-7418
2019, Vol. 6, No. 2, 167-178.
DOI: https://doi.org/10.17979/reipe.2019.6.2.5687

Mathematics and science attitude inventory: validation for use in a Spanish-speaking context

Inventario de actitudes hacia las matemáticas y las ciencias: validación para su uso en un contexto hispano hablante

Milagros Bravo-Vick, Pascua Padró-Collazo, Michelle Borrero
Universidad de Puerto Rico

Abstract

Research and evaluation focused on students' attitudes towards science and mathematics require the availability of culturally appropriate instruments in the language of the studied population. We present the translation and adaptation of the *Mathematics and Science Attitude Inventory* for its use with Puerto Rican Spanish-speaking secondary school students, within the evaluation of a teacher professional development project. We used a cross-cultural translation and adaptation model that frames these processes in the context of establishing validity and reliability of a measure by assessing the equivalence of the original version and the translated one in various dimensions: semantic, content, technical, criterion, and conceptual. Results obtained provide evidence of the equivalence between the English and Spanish versions of the inventory, as well as the reliability and validity of both versions for our context. The feasibility and utility of the cross-cultural model used were also demonstrated. This model is a valuable guide for the translation and cultural adaption of research and evaluation instruments in diverse languages and cultures.

Keywords: attitude measures, linguistic-cultural adaptation model, STEM education, student attitudes, translation and back translations

This material is based upon work supported by the National Science Foundation under Grant No. 1038166.

Milagros Bravo-Vick orcid.org/0000-0001-6891-6144: Center for Science and Math Education Research, Universidad de Puerto Rico, Recinto de Río Piedras, San Juan, Puerto Rico.

Pascua Padró-Collazo orcid.org/0000-0002-5121-1911: Center for Science and Math Education Research, College of Education, Universidad de Puerto Rico, Recinto de Río Piedras, San Juan, Puerto Rico.

Michelle Borrero orcid.org/0000-0002-9994-361X: Center for Science and Math Education Research, Department of Biology, College of Natural Sciences, Universidad de Puerto Rico, Recinto de Río Piedras, San Juan, Puerto Rico.

Correspondence related to this article: Michelle Borrero – michelle.borrero@upr.edu

Resumen
Los esfuerzos por llevar a cabo investigación y evaluación acerca de las actitudes de estudiantes hacia las ciencias y las matemáticas dependen de la disponibilidad de instrumentos culturalmente adecuados que permitan su valoración en el lenguaje de la población estudiada. Presentamos aquí la traducción y adaptación del *Mathematics and Science Attitude Inventory*, para su uso con estudiantes puertorriqueños hispanoparlantes de escuela secundaria, en el contexto de la evaluación de un proyecto de desarrollo profesional de maestros. Utilizamos un modelo de traducción y adaptación intercultural que ubica estos procesos en el contexto de establecer validez y confiabilidad de una medida mediante el avalúo de la equivalencia de la versión original y la traducida en varias dimensiones: semántica, de contenido, técnica, de criterio y conceptual. Los resultados obtenidos proveen evidencia de la equivalencia entre las versiones en inglés y español del inventario, además de evidenciar la confiabilidad y validez de ambas versiones en la población enfocada. Se demuestra, además, la viabilidad y utilidad del modelo utilizado. Dicho modelo es una guía valiosa para la traducción y adaptación de instrumentos de recopilación de datos para su uso en diversos lenguajes y culturas.

Palabras clave: medición de actitudes, modelo de adaptación lingüístico-cultural, educación STEM, actitudes de estudiantes, traducción y traducción inversa

The current importance of mathematics and science makes it imperative to understand student affective dispositions towards these subjects. Research on attitudes towards science (e.g., Osborne, Simon, & Collins, 2003) and mathematics (e.g., Aydin, 2016) is thus needed. Assessment of these attitudes require culturally appropriate instruments in the language of the studied population (Minner, Ericson, Wu, & Martinez, 2012). We faced this need in the evaluation of an educational project in Puerto Rico called Maximizing Yield through Integration (MYTI) (Borrero, Bravo-Vick, Fortis, Padró-Collazo, 2016).

An alternative to creating new instruments for research or program evaluation is to identify appropriate, already-existing instruments, even if not in the language of the target population. The MYTI project had as one of its objectives to improve student attitudes towards mathematics and science as a result of a professional development intervention with teachers. As part of our evaluation, we identified an inventory in English called the Mathematics and Science Attitude Inventory (MSAI). This inventory was created by Elizabeth Paciorek (1997), at the Rochester Institute of Technology in New York for the evaluation of Project Edge, which had a similar objective to that of MYTI and the same sponsor (US National Science Foundation).

However, adequacy of an instrument in one culture or subculture, even if in the same language, does not guarantee its validity in another (Brislin, Lonner, & Thorndike, 1973). For this reason, instrument validity and reliability need to be reestablished, especially if translation is involved (Koballa & Glynn, 2008). Therefore, we needed to translate the inventory into Spanish, assess its appropriateness for our target population, and obtain evidence of the linguistic and cultural equivalence between the original and the translated versions for our population. A translation and adaptation model, developed by Flaherty (1987) was used for this purpose.

Appropriateness (whether the instrument *should* be translated) and feasibility (whether research instruments *can* be translated) need to be addressed in translating research instruments (Behling & Law, 2000). Opposing positions on *appropriateness* are: 'translations should not be carried out because constructs are unique to each culture' versus 'it is appropriate because many constructs are similar across different cultures'. Behling and Law (2000) concluded that the transferability of constructs and measures from one culture to another must be addressed on a case-by-case basis, since some phenomena are clearly emic (limited to a single culture), while others are etic (persisting in recognizable form across cultures). Meanwhile, translation *feasibility* is restricted by the lack of: semantic equivalence across languages, conceptual equivalence across cultures, and normative equivalence across societies.

The translated version of a research instrument must satisfy various sets of requirements for useful measures, specifically: basic standards of validity, reliability, and utility for all measures, whether created or translated; acceptable levels of semantic and conceptual equivalence relative to the source language; and appropriate administration procedures that minimize problems created by the lack of normative equivalence across societies (Behling & Law, 2000).

Culturally sensitive and valid instruments and measures are needed for research and evaluation on attitudes. To enable drawing valid conclusions specifically on program evaluation, Conner (2004) emphasized the need for multicultural validity that involves three categories, two apply to the research design (*internal* and *external validity)*, while *construct validity* is pertinent to instruments that assess constructs like attitudes.

Translation and Adaptation Model

The issues and requirements mentioned above are addressed in the model we used to translate and adapt the MSAI from English. Originally, the model was posed for mental health research (Flaherty, 1987). Since attitude is an affective disposition, we consider this model to be appropriate for an attitude inventory vis-à-vis models used in the education testing literature (e.g., Hambleton, 2005). Moreover, this model has been used in Puerto Rico to translate and adapt structured research instruments for mental health (Bravo, Canino, Rubio-Stipec, & Woodbury-Fariña, 1991) and education (i.e., Bellido, Wayland, Bravo, Fortis, & Arce, 2013). In addition, the model has been used to explain the cultural adaptation of research instruments for ethnic minority research (Bravo, 2003).

The model frames the translation and cultural adaptation of research instruments in the context of establishing its validity. It can be used for simultaneous development of an instrument in various languages or for translation and adaptation of an already-existing instrument. The model postulates that the cross-language and cross-cultural equivalence of instrument versions can be established by documenting equivalence in five dimensions.

1. Content equivalence. This implies that each item evaluates a phenomenon that occurs in and is noted as real by members of the studied cultural groups. A multidisciplinary committee familiar with the instrument's constructs, and the source and/or target cultures, must judge if each item is relevant for the target culture; non-relevant items should be deleted. The translated or adapted instrument should also be examined for reliability (Flaherty, 1987).

2. Semantic equivalence. This refers to similarity in the meaning of each item in the languages of the studied groups. For an existing instrument, it requires a thorough process of accurate translation (see Behling & Law, 2000, for different translation methods). Consistent results from *testing bilinguals* in both languages also provides empirical evidence for this dimension.

3. Technical equivalence. This occurs when the measuring technique is similarly appropriate, that is, produces comparable effects, in different cultures. A bicultural committee, familiar with the target population, needs to consider the capabilities of the respondents and their familiarity with the instrument's format and administration technique. Assessing feasibility of the instrument's adapted version (evaluating whether the instrument can be administered and answered by target participants) also provides evidence for this dimension.

4. Criterion equivalence. Criterion validity measures how well one measure predicts an outcome for another measure (Linn & Gronlund, 1995). This implies that the interpretation of results obtained from the measure is similar when evaluated in accordance with each culture's established norms. In the case of a translation it involves taking into account whether the criterion used as a validator is culturally appropriate according to the new culture's norms. This involves techniques like those used to assess criterion validity of a new measure.

5. Conceptual equivalence. This indicates that the same theoretical construct is assessed in the different cultures. In the case of an existing instrument it implies that the construct or concept measured in the source instrument, exist in a similar form in the thinking of the target culture's members. Evidence of conceptual equivalence is

obtained if both versions of the instrument produce similar results when appropriate procedures are used, like those traditionally used to document construct validity. A current and more comprehensive definition of construct validity is that it involves a "determination of the significance, meaning, purpose, and use of scores from an instrument" (Creswell, 2012, p. 618). We think that the model we used includes elements consistent with those included in this definition, i.e., content (significance), semantic (meaning), technical (use), and criterion and conceptual (purpose).

Method

Participants

We administered the inventory to 60 eighth-grade students from the University of Puerto Rico, Rio Piedras Campus' laboratory-secondary school. These participants were selected based on their grade level (8th grade), since we wanted to adapt the inventory to middle school students, and also based on their science teachers' willingness to collaborate in this process. Both students and their parents or guardians signed a consent for their participation. The gender and age distributions of the participants were: 34 females and 23 males (3 missing-data), from 12 (n = 1) to 14 (n = 18) years of age, with a mode of 13 years (n = 37) (4 missing-data).

We identified a subset of students that we considered 'bilingual', since they answered at least 75% of items correctly in an English comprehension test (English Club, 2014). This subgroup (*n* = 39) was used to test the semantic equivalence between the Spanish and English versions, and assess the reliability and validity of the latter. To test the other equivalency dimensions for the translated Spanish version, we used the whole group (*n* = 59; 1 missing-data).

Instrument

The MSAI was designed to measure high-school and college students' attitudes towards mathematics and science. It includes 62 statements that describe the way students feel about these disciplines using a 4-point Likert-type scale (Strongly Disagree; Disagree; Agree & Strongly Agree); 36 items depict positive attitudes (e.g., "Solving mathematics problems is fun") and 26 negative attitudes (e.g., "I don't do very well in science"). The inventory includes items that address the three components identified in Reid's (2006) definition of attitude: Cognitive, 20 items, (e.g., "Science is of great importance to a country's development)"; Affective, 20 items (e.g., "Mathematics is something which I enjoy very much"); and Behavioral, 22 items (e.g., "Sometimes I read ahead in my science book."). Half of the statements measure attitudes towards mathematics (31 items) and half attitudes towards science (31 items); both sets are worded identically except for the discipline they address. We divided the inventory into these two discipline-content sub-inventories for administration purposes since we planned to assess middle-school and high-school students. For validity testing purposes, we classified the items into those related to school (school-related attitudes; e.g., "I would like to spend less time in school doing mathematics.") and those concerning society in relation to disciplines/professions (society-related attitudes; e.g., "Science is useful for the problems of everyday life."), a distinction reported by Osborne et al. (2003) in their literature review.

We selected this inventory because: a) our target population for evaluation was large, so we needed a survey instrument that produced quantitative results (Tytler, 2014). b) The project evaluated integrated mathematics and science; hence, we needed to evaluate attitudes towards both disciplines. c) An initial inspection indicated that the inventory was culturally appropriate for our population. d) Items consist of short, relatively simple sentences; an asset for maximum translatability (Brislin et al., 1973). e) It was developed and used in another NSF-funded project with an objective similar to ours. Our main limitation was that we could not find information about its validity and reliability in the original English version, even though we tried to contact its developer and searched pertinent literature. To address this, we obtained evidence for these psychometric properties in the process of obtaining evidence for equivalence between the original and the version translated for our context. Our aim was to have an instrument with equivalent validity and reliability for our target population for both the original English and the translated Spanish versions.

Procedure

The MSAI was administrated to the students by their science teacher during a 50 minute science class time-period. Details of the procedures used to obtain equivalence in each of the model's five dimensions are described below.

Content equivalence. This was assessed in two ways. First, judging the relevance and cultural appropriateness of each inventory item for measuring attitudes towards mathematics and science in our target population, that is, whether the attitude constructs operationalized in the items exist in our context in a form similar to that described in the original version. This qualitative procedure, similar to that used to assess content validity of new instruments, was carried out by four Puerto Rican bilinguals: one STEM intermediate-school teacher, one STEM college professor and two evaluators described below. Second, measuring the internal consistency (reliability evidence) of the inventory's Spanish translated-version and of the original English version in our context.

Semantic equivalence. Two processes were used. First, translation/back translation, which consists of translating the original instrument from English to Spanish, back translating it to English and then comparing the two English versions (Behling & Law, 2000). Each step was carried out by one of three bilingual persons working independently. One evaluator, a Ph.D. education psychologist, carried-out the English-to-Spanish translation. Another evaluator, Ed.D. in science education who studied in US schools for 4-years during her formative years, carried out the back-translation from Spanish-to-English. A translation graduate student compared the original English version and the English back-translation; she checked whether each item had a similar meaning/intent in both languages (the wording does not have to be the same). Second, the administration to bilingual students in English and Spanish with a time lapse of at least seven days, with half of the students receiving the English version and half the Spanish version first (counter-balanced), to control for administration order bias.

Technical equivalence. Various processes were used. The two evaluators and an intermediate school science teacher evaluated whether the target population could manage adequately the instrument's format and technique. Then the instrument was divided into two parts ('attitudes to mathematics' and 'attitudes to science') for administration since we thought that the original instrument, designed for high-school and college students, was too long for middle-school students. Bias due to administration order was avoided by counter-balancing the administration order of the sub-inventories. Finally, the verbal reports from teachers who administered the inventory and observations from one of the evaluators who was present during the first set of administrations were used to determine the ease of use of both versions.

Criterion equivalence. A process similar to that used to assess concurrent validity (whether the instrument evaluates the construct/concept it is supposed to assess) was used. An established valid instrument is usually used as criterion, but we were not able to identify a pertinent valid instrument (used with a Puerto Rican population in either English or Spanish). However, scores of a mathematics and science attitude inventory might be expected to be concurrent with student preference for their science or math class, since preference, like attitude, is an emotional disposition toward an object. Thus, preference could be used as a criterion. Moreover, in our cultural context, asking students about this preference is a generalized practice and therefore appropriate. We thus expected a strong relationship between preference and attitude, and a similar relationship in both the English and Spanish versions.

Conceptual equivalence. To assess this equivalence, we used methods employed to test construct-related validity of created instruments.

For convergent validation we hypothesized that school grades (sum of numerical values of grades in 7th & 8th grade) would be positively associated with mathematics and science attitudes (based on, e.g., Chow, Eccles, &

Salmela-Aro, 2012). We also expected that correlations with school-related attitudes would be higher than with society-related attitudes since grades 'exist' in the school context. For *divergent validation* we hypothesized that students who expressed preference for both STEM-related college studies and careers would show significantly higher levels of positive attitudes towards mathematics and science than those who did not express these preferences, as documented in multiple studies (e.g., Engberg & Wolniak, 2013). Difference between groups would be larger for society-related than school-related attitudes since 'college studies' and 'careers' are indicators of a professional role played in society.

To obtain evidence based on internal structure, we carried out an exploratory factor analysis (Maximum Likelihood Estimation). The factorial solution was obtained independently for the mathematics and science sub-inventories, to meet the requirement that the matrix of the original data has a larger number of cases than of variables. Since the sample size was less than 100 for both language versions, comparisons were mainly limited to the value of coefficients corresponding to items extracted for the first factor. The correlation between the coefficients of items (in the Spanish and English versions) in the first extracted factor was estimated to evaluate the possible likelihood of the factor structure of both language versions.

Results

Results are summarized for the entire set of items on the MSAI, subdivision by mathematics or science attitudes, and subdivision by 'school-related' and 'society-related' items.

Semantic Equivalence

Translation/Back translation process. In the comparison of the MSAI's original English version with the English back-translation two items did not meet the requirement of similarity in the items' intent/meaning (out of 31 in the science sub-inventory; both sub-inventories are worded exactly the same with the exception of the subject matter). These items were reworded in consultation with the translator. That only two items were identified is probably due to the relatively simple language and short sentences of the inventory**.**

Administration to bilinguals in both languages. High and significant levels (p < .001) of inter-language correlations were observed for the entire MSAI (r = .91), and also for the mathematics (r = .87) and science (r = .81) sub-inventories (n = 39). These results provide additional evidence for semantic equivalence between the English and Spanish versions of the inventory.

Content Equivalence

Review by bilinguals. The bilingual teachers and evaluators who reviewed the inventory indicated that all its items are appropriate to assess attitudes towards mathematics and science in the new language (Spanish) and context (secondary school students in Puerto Rico). That is, the inventory content is noted as real and culturally appropriate for the target population.

Reliability of the Spanish and English inventories. The English and Spanish versions of the inventory showed similarly high levels of internal consistency (Table 1) for the entire set of items on the whole MSAI and the mathematics and science sub-inventories (α > .90). These results also provide evidence for content equivalence between the two language versions and for the reliability of each of these versions in our context.

Technical Equivalence

Ease of usage of the Spanish and English versions. The teachers who administered the inventory, and the evaluator who observed the first administration in each language, reported that no problems were identified in the administration or the students' answering of the inventory in either English or Spanish.

Table 1
Internal Consistency (Cronbach's Alpha Coefficient) of the Inventory's English (n = 39) and Spanish (n = 59) Versions

Item classification	Version	Set of items		
		School-related	Society-related	Entire set
Mathematics and Science Inventory	English	.90	.86	.93
	Spanish	.88	.88	.93
Mathematics Sub-Inventory	English	.90	.85	.93
	Spanish	.85	.84	.91
Science Sub-Inventory	English	.88	.85	.92
	Spanish	.85	.83	.90

Criterion Equivalence

Relationship of inventory scores with a criterion. The hypothesis that there would be a strong positive relationship between preference for a class and attitude to the corresponding subject matter was confirmed in our study. Students ranked the classes they were taking in order of preference with the highest number (7) being assigned to their first preference and the lowest number (1) to their last. Results show that the higher the student preference for mathematics/science class, the higher the score for the corresponding attitudes sub-inventory. The pertinent correlations were all positive, significant ($p \leq .001$), and large (L) (Cohen, 1998) in both languages, although larger for the original English version (Table 2).

Table 2
Relationship (Pearson's r) between Mathematics/Science Class Preference and Corresponding Attitudes

Hypothesis	Subject	English (n = 39)	Spanish (n = 59)
		Results (Interpretation)	
The higher student *preference for mathematics/science class*, the more positive *the corresponding attitudes.*	Math	.75* (L)	.55* (L)
	Science	.73* (L)	.57* (L)
The higher the student *preference for mathematics/science class*, the more positive the corresponding *school-related attitudes.*	Math	.70* (L)	.52* (L)
	Science	.71* (L)	.60* (L)
The higher the student *preference for mathematics/science class*, the more positive the *corresponding society-related attitudes*.	Math	.69* (L)	.53* (L)

*$p \leq .001$

Conceptual Equivalence

Convergent validation. We hypothesized that the higher the students' school grades in mathematics/science, the stronger the positive relationship with attitudes toward the corresponding discipline. This value, ranging from 1 to 8, was a result of the sum of each students' 7th and 8th grade science and mathematics final course grades. This hypothesis was confirmed for mathematics ($p \leq .01$ or $p \leq .001$) in both language versions (Table 3, row 1), except for the Spanish version's society-related attitudes to mathematics (row 3); correlations were stronger (S = Small, M = Medium, and L = Large; Cohen, 1998) for the English version. The relationships for science attitudes were not statistically significant.

The hypothesis that these correlations would be higher for school-related attitudes than for society-related ones was confirmed (Table 3, rows 2 & 3), but only for the Spanish version regarding mathematics: moderate and significant for school-related attitudes, and low and non-significant for society-related ones. Contrary to expectations, all the other corresponding comparisons were of similar magnitude.

Table 3

Relationship (Pearson's r) between Students' Mathematics/Science (M/S) Grades and Attitudes to M/S Inventory Scores

Hypothesis	Discipline	English (n = 39)	Spanish (n = 59)
		Results (Interpretation)	
The higher the grades in mathematics/science the more positive the attitudes towards the corresponding discipline.	Math	.55*(L)	.32*(M)
	Science	.27 (S)	.18 (S)
The higher the grades in mathematics/science, the more positive the *school-related* attitudes towards the corresponding discipline.	Math	.51**(L)	.40*(M)
	Science	.25 (S)	.22 (S)
The higher the grades in mathematics/science, the more positive the *society-related attitudes* towards the corresponding discipline.	Math	.52**(L)	.20 (S)
	Science	.26 (S)	.12 (S)

*$p \leq .01$. **$p \leq .001$

Divergent validation. We hypothesized that students who indicated they would like to both study a STEM discipline and practice a STEM-related profession in the future (Group 1) would show significantly higher levels of positive attitudes towards mathematics and science than students who mentioned either a STEM discipline or career or none (Group 2). As hypothesized, in both the English and Spanish versions, Group 1 students showed significantly ($p \leq .01$) better attitudes to mathematics (Table 4, row 1), and to society-related mathematics attitudes (row 3) than Group 2 ($p \leq .05$) students. However, regarding school-related mathematics attitudes (Table 4, row 2), the hypothesis was confirmed for the English version but not the Spanish one. None of the hypotheses regarding science attitudes were confirmed.

Table 4

Hypothesized differences in attitudes to mathematics/science between students interested (Group 1) or not (Group 2) in STEM studies and careers

Hypothesis	Subject	English Version (n = 39) Group 1 (n = 25) Mean (SD)	English Version Group 2 (n = 14) Mean (SD)	English Version U Mann Whitney test	Spanish Version (n = 58) Group 1 (n = 37) Mean (SD)	Spanish Version Group 2 (n = 21) Mean (SD)	Spanish Version U Mann Whitney test
1. Group 1 will show more positive *attitudes towards math or science* than Group 2.	Math	88.4 (14.2)	98.9 (10.4)	2.519**	91.7 (13.1)	98.3 (11.8)	1.960*
	Sci.	90.7 (12.5)	92.5 (12.0)	0.557	93.7 (12.9)	91.4 (10.6)	0.688
2. Group 1 will show more positive *school–related attitudes towards math or science* than Group 2.	Math	44.5 (8.2)	50.2 (6.4)	2.250*	46.0 (6.9)	48.6 (6.8)	1.443
	Sci.	45.7 (7.5)	46.6 (6.9)	0.323	46.9 (6.7)	45.0 (6.3)	0.892
3. Group 1 will show more positive *society- related attitudes towards math or science* than Group 2.	Math	43.9 (7.1)	48.6 (4.9)	2.113*	45.7 (7.5)	49.8 (5.5)	2.260*
	Sci.	45.0 (5.5)	45.9 (6.1)	0.499	46.8 (6.9)	46.4 (5.3)	0.389

*$p \leq .05$, **$p \leq .001$

The hypothesis that the differences between groups would be larger for society-related than school-related attitudes (Table 4, rows 2 & 3) towards mathematics was confirmed for the Spanish version (differences were larger and significant for society-related attitudes, and smaller/non-significant for school-related ones), but not for the English version (z values were both significant and similar in magnitude). Contrary to expectations, the corresponding hypothesized comparisons for science attitudes were not confirmed.

Factor analysis. Results of the exploratory factor analysis indicate that the first factor explained a high percentage of variance of both sub-inventories (mathematics and science; 31 items each) in both languages (29% or more) (Table 5). Also, the Pearson correlations between coefficients for the first factor of both sub-inventories are large (L) (Cohen, 1988) and highly significant ($p \leq .001$). In all cases, the eight-factor solution explained 80% of the total variance (data not shown). These results suggest that the factorial structure observed for the English and Spanish versions of both sub-inventories are very similar, but these results are considered preliminary due to sample size (n = 39 for the English version and n = 59 for the Spanish version). Results thus provide additional evidence of the conceptual equivalence among the MSAI's original and the translated version. This and other results previously presented also provide evidence of the construct validity of each inventory's language version.

Table 5
Results concerning the first factor of the exploratory factor analysis

		n	Eigenvalues	% of Variance Explained	Pearson correlation (*n* = 31)
Mathematics Sub-inventory	English	39	10.8	34.8	.78*(L)
	Spanish	59	9.4	30.3	
Science Sub-inventory	English	39	10.9	35.2	.60*(L)
	Spanish	59	8.9	29.0	

*$p \leq .001$

Discussion

We obtained good evidence for the linguistic and cultural equivalence of our MSAI Spanish version and the original English version in our context. Good results were obtained for processes involving expert judgment, used to evaluate *semantic, content* and *technical equivalence;* as well as processes that required testing, used to examine *semantic, content, criterion* and *conceptual equivalence*.

Evidence for *semantic equivalence* between the MSAI's Spanish and English versions included good results from the back-translation process, and a strong relationship between scores in the English and Spanish versions. Evidence for *content equivalence* between versions included conclusions of bilingual professionals' analysis indicating that item content is relevant and appropriate for our context, and high internal consistency/reliability results for both language versions (much higher than .65, considered appropriate by Flaherty, 1987). Evidence for *technical equivalence* included professionals' judgements that the target population could adequately manage the instrument's format and technique, and reports regarding no difficulties in students answering either version, based on observations. Evidence for *criterion equivalence* included similar significant, positive, and mostly large correlations between inventory scores and the concurrent criterion (preference for the corresponding mathematics/science class) for both language versions. Evidence for *conceptual equivalence* comprised assessment of the inventories' convergent (hypothesized positive relation between attitudes and school grades) and divergent validation (hypothesized difference in attitudes between discrepant groups regarding STEM-related study and work variables) and factor analysis. The corresponding hypothesized relations were confirmed for mathematics for both language versions but not for science. However, the exploratory factor analysis provided evidence of the conceptual equivalence between the language versions for both the mathematics and science sub-inventories. Moreover, the applicability of the translation/adaptation model used was evidenced by the processes used and results obtained to document interlanguage/cultural equivalence in this study.

In the process of testing the previously mentioned equivalence dimensions, we also obtained evidence for *feasibility, reliability,* and *validity* for both language versions in our context. Evidence of construct validity, as currently defined (Creswell, 2012), was also obtained for both MSAI's language versions while documenting the five equivalence dimensions: determination of significance (content), meaning (sematic), purpose (criterion and conceptual) and use (technical) of scores from an instrument.

A puzzling finding was the tendency of some quantitative equivalence results (i.e., criterion and construct) to be consistently higher for the English than for the Spanish version, even though similar very high results were observed for others (i.e., semantic and technical). One can speculate that although the Spanish translation/adaptation was relatively good it was not as good as the original English version. However, the positive results observed for mathematics on all equivalence dimensions in both languages, and the similar non-significant results for both language versions for the science sub-inventory regarding conceptual equivalence, suggest that the wording of items in Spanish was probably not the problem. The non-significant results observed for science on conceptual equivalence (convergent-divergent validation) are also puzzling because we consider that the validators used (school grades and intention to study and work in STEM-related careers), that produced significant results for mathematics attitudes, are similarly appropriate for science. A plausible explanation is that administration by two science teachers in the science class might have influenced student reports of attitudes towards science or towards a validation measure (i.e., STEM college studies/careers).

This study addresses the need for greater emphasis on understanding affective elements regarding science/mathematics learning (Niss, 2007; Osborne et al., 2003). First, it documented inter-language/cultural equivalence between the MSAI original English version and a Spanish translation for secondary school students in Puerto Rico. Second, it provided evidence for the reliability and validity of the MSAI in its original English version in our context. However, this psychometric evidence may not apply to English speakers in other contexts, especially to native speakers. When an instrument is to be used in a different population, reliability and validity need to be reestablished (Koballa & Glynn 2008). Third, the study provided evidence for the reliability and validity of a Spanish version of the MSAI for our context. Fourth, it illustrated the feasibility and utility of a model addressing issues identified as important for the translation/adaptation of research instruments for diverse linguistic and cultural contexts (Behling & Law, 2000; Conner, 2004). Thus, this model can be used to carry out careful linguistic and cultural adaptations, including adaptation of our Spanish-translated MSAI for other Spanish-speaking contexts. Fifth, the two versions of the MSAI can be used in inter-language and cross-cultural studies, after appropriate linguistic and cultural adaptations for multicultural validity. Various international studies have identified differences across cultures in student attitudes towards mathematics or science (e.g., Zhao & Singh, 2011), thus more research is needed to understand them.

This study had several limitations. First, there was potential sampling bias, because it was limited to 8th grade students from one school in the San Juan metropolitan area. Second, there was potential response bias since the instrument is a self-report survey that permits dishonest responses. Although the item content was not sensitive, we tried to minimize this by assuring confidentiality to participants both in written and oral statements. A third limitation was that the bilinguals who tested the English version were part of the sample that tested the Spanish version. For testing semantic equivalence, the use of the same participants is necessary, for testing other equivalences (especially criterion and conceptual), the use of different samples provides more independent results. Also, the reliability test used for internal consistency was not as robust as other methods for determining the reliability of scores, such as test-retest and alternate forms; internal consistency is the degree to which instrument items combine to measure a single trait (Henson, 2001). However, the administration to the same bilingual students in both languages could be considered a kind of test-retest, and we obtained high levels of inter-language correlations (e.g., $\alpha > .9$ for the whole MSAI) using this procedure. Finally, an exploratory factor analysis (focusing mainly on results of the first factor, due to small sample size) was used to get additional evidence (besides that of convergent/divergent validation) to assess conceptual inter-language equivalence and construct validity. A larger sample would have enabled carrying out a factor analysis that produced more robust results about the language versions' factor structure and its inter-language equivalence. These limitations occurred mainly because the inventory was tested and used in the context of a project evaluation, not in a research project with funds especially allocated for instrument development.

Translation and adaptation of instruments developed in one culture for use in another can enhance research and evaluation in STEM education. Instruments developed in one language and culture can be used in other contexts, after proper translation and adaptation. Findings from one setting can be corroborated in diverse contexts, augmenting evidence of external validity, or generalizability to other groups over different times and settings. Evaluation of outcomes of specific STEM education interventions can also be studied in diverse contexts while recognizing the diversity of phenomena across cultures and subcultures (Thompson-Robinson, Hopson, & Sen-Gupta, 2004). In our case, it enabled the evaluation of an intervention, specifically, the impact of a STEM education project on the attitudes of 6th -12th grade students from schools located throughout Puerto Rico. We plan to publish these results soon and to continue studying attitudes in a future project.

Thus, the present study provided evidence of reliability and validity of the MSAI for the original English inventory and the translated Spanish version in our context. These versions can be used in research and evaluation in each of these languages in other populations, after appropriate adaptations and testing. More importantly, the study provided evidence for equivalence between the English and Spanish versions in our context, enabling their use for the evaluation of a STEM project at our location. The model used is a valuable guide for the translation and cultural adaption of research and evaluation instruments in diverse languages and cultures.

References

Aydin, B. (2016). Examination of the relationship between eighth grade students' learning styles and attitudes towards mathematics. *Journal of Education and Training Studies, 4*(2), 124-130. https://doi.org/10.11114/jets.v4i2.1162

Behling, O., & Law, K.S. (2000). *Translating questionnaires and other research instruments: Problems and solutions*. Thousand Oaks, CA: SAGE. https://doi.org/10.4135/9781412986373

Bellido, C., Wayland, K., Bravo Vick, M., Fortis, M. & Arce, J. (2013). *Manual de Uso del Protocolo de Observación en la Educación Reformada (PROEDUCAR)* [Usage Manual of the Observation Protocol for Reformed Education]. Retrieved from: http://alacima.uprrp.edu/Nuevo-Portal/?page_id=27

Borrero, M., Bravo-Vick, M., Fortis, M. & Padró-Collazo, P. (2016). Mejorando la educación STEM a través de la investigación educativa [Improving STEM education by means of educational research]. *Acta Científica 30* (1-3), 6-25.

Bravo, M. (2003). Instrument development: Cultural adaptations for ethnic minority research. In G. Bernal, J. E. Trimble, A. K. Burlew & F. T. L. Leong, *Handbook of Racial and Ethnic Minority Psychology.* Thousand Oaks, CA: Sage.

Bravo, M., Canino, G. J., Rubio-Stipec, M. & Woodbury-Fariña, M. (1991). A cross-cultural adaptation of a psychiatric epidemiologic instrument: The diagnostic interview schedule's adaptation in Puerto Rico. *Culture, Medicine and Psychiatry, 15*(1), 1-18. https://doi.org/10.1007/BF00050825

Brislin, R. W., Lonner, W., & Thorndike, R. M. (1973). *Cross-cultural research methods.* New York: John Wiley. https://doi.org/10.1177/136346157501200101

Chow, A., Eccles, J.S., & Salmela-Aro, K. (2012). Task value profiles across subjects and aspirations to physical and IT-related sciences in the United States and Finland. *Developmental Psychology 48*(6), 1612-1628. https://doi.org/10.1037/a0030194

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York, NY: Psychology Press. https://doi.org/10.1016/C2013-0-10517-X

Conner, R. F. (2004). Developing and implementing culturally competent evaluation: A discussion of multicultural validity in two HIV prevention programs for Latinos. In M. Thompson-Robinson, R. Hopson, & S. Sen-Gupta (Eds.), *In search of cultural competence in evaluation: Toward principles and practices*. New Directions in Evaluation, 102, 51-65. San Francisco, CA: Jossey-Bass. https://doi.org/10.1002/ev.115

Creswell, J. W. (2012). *Educational Research*. Boston, MA: Pearson.

Engberg, M. E., & Wolniak, G. C. (2013). College students' pathways to STEM disciplines. *Teachers College Record*, *115*(1), 1-27. https://www.tcrecord.org/content.asp?contentid=16740

English Club (2014). *Test of Written Comprehension* [Measurement instrument]. Retrieved from http://www.englishclub.com/esl-exams/levels-test-wc-start.htm.

Flaherty, J. A. (1987). Appropriate and inappropriate research methodologies for Hispanic mental health. In M. Gaviria (Ed.), *Health and behavior: Research agenda for Hispanics* (pp. 177-186). Chicago: University of Illinois Press.

Hambleton, R. K. (2005). Issues, designs and technical guidelines for adapting tests into multiple languages and cultures. In R. K. Hambleton, P. F. Merenda, C. D. Spielberger (Eds.), *Adapting Educational and Psychological Tests for Cross-Cultural Assessment* (pp. 3-38). New Jersey: Lawrence Erlbaum Associates, Inc., Publishers. https://doi.org/10.4324/9781410611758

Henson, R. K. (2001, January 26). *Teacher self-efficacy: Substantive implications and measurement dilemmas*. Paper presented at the Annual Meeting of the Educational Research Exchange. College Station, TX: Texas A&M University. Retrieved from https://files.eric.ed.gov/fulltext/ED452208.pdf

Koballa Jr., R., & Glynn, S. M. (2008). Attitudinal and motivational constructs in science learning. In S. K. Abell, & N. G. Leaderman (Eds.), *Handbook of research in science education* (pp. 75-102). New York: Routledge Taylor & Francis Group. https://doi.org/10.4324/9780203824696

Linn, R. L., & Gronlund, N. E. (1995). *Measurement and assessment in teaching* (7th ed.). New Jersey: Prentice-Hall, Inc.

Minner, D., Ericson, E., Wu, S., & Martinez, A. (2012). *Compendium of STEM student Instruments PART II: Measuring students' content knowledge, reasoning skills, and psychological attributes*. Washington, DC: Abt Associates. Retrieved from http://cadrek12.org/resources/compendium-research-instruments-stem-education-part-ii-measuring-students-content-knowledg

Niss, M. A. (2007). Reflections on the state and trends in research on mathematics teaching and learning: From here to utopia. In F. K. Lester (Ed.), *Second Handbook of Research on Mathematics Teaching and Learning* (Vol. 2, pp. 1293-1312). Charlotte, North Carolina: Information Age Publishing, Inc.

Osborne, J., Simon, S., & Collins, S. (2003). Attitudes towards science: A review of the literature and its implications. *International Journal of Science Education*, *25*, 1049-1079. https://doi.org/10.1080/0950069032000032199

Paciorek, E. (1997). *Mathematics and Science Attitude Inventory* [Measurement instrument]. Rochester Institute of Technology "Project EDGE." Retrieved from http://oerl.sri.com/instruments/up/studsurv/instr129.html

Reid, N. (2006). Thoughts on attitude measurement. *Research in Science & Technological Education*, *24*(1), 3-27. https://doi.org/10.1080/02635140500485332

Thompson-Robinson, M., Hopson, R., & Sen-Gupta, S. (2004). In search of cultural competence in evaluation: Toward principles and practices. *New Directions in Evaluation*, *102*. San Francisco, CA: Jossey-Bass.

Tytler, R. (2014). Attitudes, identity, and aspirations towards science. In N. G. Lederman, & S. K. Abell (Eds.), *Handbook of Research on Science Education* (Vol II, pp. 82-103). New York, NY: Routledge. https://doi.org/10.4324/9780203097267

Zhao, D. & Singh, M. (2011). Why do Chinese-Australian students outperform their Australian peers in mathematics: A comparative case study. *International Journal of Science and Mathematics Education 9*, 69-87. https://doi.org/10.1007/s10763-010-9214-7

Fecha de recepción: 2 de agosto de 2019.
Fecha de revisión: 25 de noviembre de 2019.
Fecha de aceptación: 26 de noviembre de 2019.
Fecha de publicación: 1 de diciembre de 2019.