

International Conference on Education and Educational Psychology (ICEEPSY 2012)

Bridging the gap between conceptualisation & assessment of intercultural competence

Joana Almeida^{a*}, Ana Raquel Simões^a, Nilza Costa^a

^aResearch centre Didactics and Technology in Education of Trainers (CIDTFF),
Department of Education, University of Aveiro, Aveiro 3800-193 Portugal

Abstract

This paper aims at deconstructing the intercultural competence assessment task by building a coherent framework to assess the IC of incoming *Campus Europae* students at the University of Aveiro in 2011/12.

Our discussion focuses on pre/post-test design survey techniques, by reflecting upon the process of moving from the first to the improved final version of a pre-test questionnaire partly adapted from the YOGA format Assessment of Intercultural Competence questionnaire.

Findings provide practical ideas on how to lessen the social desirability bias of this type of questionnaires, whilst prioritising the specific aspects of intercultural competence to be measured.

© 2012 The Authors. Published by Elsevier Ltd. Open access under [CC BY-NC-ND license](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Selection and peer-review under responsibility of Dr. Zafer Bekirogullari of Cognitive – Counselling, Research & Conference Services C-crcs.

Keywords: intercultural assessment; intercultural instruments; intercultural competence; organised student mobility; Campus Europae

1. Introduction

The study discussed in this paper is embedded in an ongoing PhD case study¹ which aims at assessing the intercultural competence (IC) of incoming *Campus Europae* (CE) students at the University of Aveiro (UA) in 2011/12 so as to foster their intercultural learning while abroad. The overall purpose of this paper is to shed some light on the IC assessment task when it incorporates pre/post-test design survey techniques. For that reason we will map feasible answers to the challenges raised during the first phase of a tripartite assessment task, which precedes a quasi-experimental intervention plan of 8 two-hour interrelated modules given by the researcher to the target population to facilitate their IC whilst in-country.

The first phase challenges stem from the process of moving from the first version of a pre-test questionnaire to the improved final one. First administered to incoming CE students in 2010/11, this quantitative instrument is partly adapted from the YOGA format Assessment of Intercultural Competence (AIC) questionnaire², and adopts the same working definition of IC - the “complex abilities that are required to perform effectively and appropriately when

* Corresponding author. Tel.: 00351 234 372 442. E-mail address: jalmeida@ua.pt.

¹ Sponsored by the Portuguese Foundation for Science and Technology, Operational Programme Human Potential (POPH)

² Available online at www.worldlearning.org

interacting with others who are linguistically and culturally different from self” (Fantini, 2009, p. 458). This complex set includes: various attributes, three areas, four dimensions (positive attitudes, critical awareness, knowledge and skills), target language proficiency, and developmental levels (Fantini, 2006, 2009).

It is the subset of the four dimensions which was adapted to the Portuguese exchange context and to the profile of our target population. Thus, it is the difficulties arising from that adaptation that are the object of analysis in this paper. Those difficulties are here addressed through two overarching research questions:

1. *How to decrease the social desirability bias (SDB) of the questionnaire?*
2. *How to prioritise the specific aspects of IC measured by the questionnaire?*

The answers to these questions may provide valuable insights to ongoing research efforts in bridging the gap between IC conceptualisation and its assessment.

2. Conceptualising the contexts:

Although the primary focus of this study is empirical and data analysis-based, it is rooted in important theoretical frameworks: the European student mobility and the intercultural assessment contexts. It is precisely at the intersection of these two frameworks that our research problem lies, but it is also at this very intersection that the solution can be found. A depiction of these contexts will thus constitute the starting point of our discussion.

2.1. *The European student mobility context*

Today’s global flows are breaking down the barriers of geographic space, fostering more than ever before diversified everyday contacts, where the Other is no longer distant. This closeness to the Other meant, in the European educational-political field, a significant investment in student mobility posited as the flagship of the European identity.

It is therefore no surprise that one of the core action lines of the Bologna Process was the creation of a European Higher Education Area (EHEA), grounded on the convergence, competitiveness and internationalization of European Higher Education systems (Papatsiba, 2006). And yet, if this meant a significant boost in European student flows, it also led to an emphasis on mobility numbers in the European Commission political discourse. Evidence for this can be found, for instance, in the astounding funding levels of “Erasmus for all”, the new European programme for Education, Training, Youth and Sport for the period 2014-2020.

The immediate consequence of this ill-advised optimism is to cite greater output numbers as evidence of successful intercultural outcomes (Deardorff, 2006). Shifting the emphasis from the promotion of student mobility to the processes underpinning its intercultural gains becomes of paramount importance. In this particular study we attempt to do so by addressing European horizontally-organized student mobility, that is, mobility encouraged by educational programmes taking place within a cycle of studies in order to attain credits as recognition of the students’ studies abroad (Szarka, 2003; Wätcher, 2010).

In this research, this type of mobility is embodied by CE, an enhanced mobility network for students and staff within 20 universities across 16 European countries (Moreira & Amorim, 2010). Set up in 2001, under the umbrella organization European University Foundation, this network aims at fostering the ideals of a European identity and citizenship through mutual cooperation among member institutions (CE, www.campuseuropae.org). To accomplish this, CE offers a minimum of 1 and a maximum of 2 academic years abroad in two different universities/countries, believing that only by empowering students with language skills can the European ideals be achieved. This “2 years, 2 languages” concept is based upon a three-step language approach: an introductory online course of the target language via the language gateway *Hook-Up!*, followed by a 4-6 week intensive course before the start of the academic year, and language courses during the year abroad (ibid.). At the end of the year students are expected to reach the B1 level of competence, in accordance with the Common European Framework of Reference for Languages.

Believing in the potential of this structured-form of student mobility, our perspective is that a key element is still missing to stimulate the full potential of CE. The missing element is an interventionist approach, based on the assumption that “students learn effectively abroad when we intervene in their learning” (Vande Berg & Paige, 2009, p. 433) and not simply by living in a foreign country.

Such an assumption provides the conceptual foundations for the PhD of which this study is a part, and its purpose of offering a solid starting point for constructing intercultural pedagogy(ies) to facilitate the IC of CE movers. Completing such a goal implies tackling numerous issues, from course design to target population needs and training context, resources, training methods, language of instruction, assessment, instrumentation, to name just a few. Given that when cross-examining these issues the major difficulties were assessment and instrumentation-related, writing a paper about this decision-making process became more than an option but a task that could provide valuable insights to other studies committed to intercultural assessment using intercultural instruments.

2.2. *The intercultural assessment context*

The array of theoretical impulses of the cultural turn have caused an undeniable shift in language teaching and learning, introducing both the concept of IC into the language curriculum, but also causing divergence amongst methods of language teaching and assessment. Hence the often dichotomous division between the conceptualisation of IC and its assessment (Bhawuk, 1998, 2001; Van de Vijver & Leung, 2009).

In light of this controversy, both scholars and language professionals are faced with the challenge of choosing from the conceptual frameworks of IC those which can be boosted by intercultural training that can subsequently be assessed. Variety seems not to be the problem, but rather making sense of it by defining clear and explicit assessment goals. *Do all these professionals choose frameworks which are in line with the outcomes they intend to assess? Is there a real parallel between the goals of instruction and those of evaluation* (Fantini, 2009)?

Choosing an appropriate conceptual and assessment framework is a sensitive task, since a misleading identification of the subject matter, goals and objectives can compromise the reliability and quality of the assessment provided. The choice can be even harder if one chooses to use intercultural instruments as part of any sort of instructional course committed with intercultural learning. For the purpose of this study, intercultural instruments are understood as "any measurement device that identifies, describes, assesses, categorizes or evaluates the cultural characteristics of individuals, groups, and organizations" (Paige, 2004, p. 86).

Notwithstanding the myriad of intercultural measurement devices developed so far "little has been written about these instruments as a component of intercultural education or intercultural training pedagogy" (Paige, 2004, p. 85). Needless to say that leaving such an important question behind may only foster vagaries in this field of study. For this reason, authors such as Reddin (1944), Brow and Knight (1999), Fowler and Mumford (1999), Paige (2004), Dearthoff (2004, 2005) and Fantini (2009) have proposed rationales and criteria for making a thoughtful choice when selecting instruments with potential value for research.

Regardless of the criteria followed, our belief is that when incorporating external instruments in intercultural training "it is important to understand what the [chosen] instrument measures and to be sure that its purpose is compatible with the goals and objectives being assessed" (Fantini, 2009, p. 465). It is this rationale that brings about the current paper, in an attempt to take one step forward in the process of choosing and adapting intercultural instruments to the overall mission of any targeted intervention. It is not our intention to make a survey of the numerous existing instruments, but to present practical ideas on the decision-making process of using and adapting this sort of instrument across cultures, through the careful examination of the adaptation of the AIC to the Portuguese cultural context. It is precisely the decisions and ideas involved in this process that we will discuss next.

3. The study

3.1. *The intercultural instrument: selection and adaptation*

Our review of the literature revealed a daunting number of intercultural instruments and, with it, the dilemma of narrowing them down. In this process we had in mind both the extensive and selective list of 90 and 44 instruments provided by Fantini in (2006) and (2009) respectively. Narrowing down these instruments was a difficult task, aided by the establishment of 7 criteria which confined us to 3 options: the Intercultural Competence Assessment (INCA) project questionnaires, the Intercultural Development Inventory (IDI) and the Assessment of Intercultural Competence (AIC) – Yoga format.

Within these 3 options we eliminated 2 instruments, since they did not meet 2 of the 7 criteria. In the case of the INCA instrument(s) the target public was not appropriate for our study (criteria 7); as for the IDI, despite being one

of the most sophisticated and extensive tools (and of wide use), its logistic aspects were incompatible with our research timeline (criteria 6).

The fulfilment of all 7 criteria led us to select the AIC – Yoga format as the most appropriate tool. The criteria are labelled in Table 1, along with a brief explanation on how the AIC meets them.

Table 1 – Criteria in instrument selection

<u>Criteria</u>	<u>Description of AIC applicability to our study</u>
1. Construct validity	The latent variable model of IC offered by the AIC focuses both on linguistic and cultural indicators, according to 4 complementary dimensions + language proficiency
2. Type of assessment	The self-assessment feature fulfills our need to diagnose the target population's stage of development of IC previously to the intervention plan
3. Theory-grounded instrument	The theoretical support the AIC offers is theory-grounded and provides a useful frame of reference for our empirical study and a reference point for the psychometric data analysis
4. Response scale format	The several levels of the response format reflect a gradation of IC developmental process within a range of 6 competency levels (0= no competence; 5= very high competence)
5. Reliability and validity of the measurement	The reliability and validity of the AIC measurement is based on the extended form of this questionnaire which was tested with audiences of different cultural backgrounds
6. Feasibility	The instrument is cost-free and easily accessed under the time limits of this PhD project
7. Target population & development levels	The definition of 4 developmental levels corresponds to different target populations, including the 2 development levels (out of 4) that match our target population needs: Educational Traveler (I) and Sojourner (II)

After having obtained the author's permission to use and adapt this instrument the next step was to tailor its design to our research context and overall mission. The first changes made were not substantial, consisting of slight linguistic simplification and the addition of 3 items (1 extra item in the "attitudes" and 2 in the "knowledge" dimensions) to the item pool of the AIC. The purpose was to simplify the level of English used and to include items that could provide additional useful information.

In addition to these minor changes, two major changes were made: **1)** the exclusion of target language proficiency as one of the 5 dimensions of IC and **2)** the modification of the measurement format to a Likert scale (maintaining the formal 6-point scale). The first change is justified by the fact that target language proficiency was a subsidiary goal of our research, as the 8 intervention-modules were part of the Portuguese as Foreign Language (PFL) class, where the main teacher was responsible for focusing on the Portuguese language. Though the language of instruction of the intervention plan was Portuguese, our commitment was not to instigate the participants' Portuguese language proficiency, even if that could emerge as an add-on. As for the second major change, the rationale was that providing labels to the 6 degrees of competence would give respondents a clearer reference point and make interpretation easier.

Once the adaptation process was complete, a small-scale study followed aimed at testing the reliability and validity of the adapted part of the questionnaire, as well as of the sections of the instrument which were new.

3.2. The pilot study instrument: characterization and data analysis

3.2.1. Participants

The pilot instrument was administered in May 2011 to 30 of the 31 incoming CE students at UA in 2010/11, due to the impossibility of selecting a subset of our target population (incoming CE students 2011/12) later on.

Sixteen (53.3%) of these students were female and fourteen (46.7%) were male, with ages ranging between 21 and 25; the average age being 22.2 (SD= 1.2). Most students were frequenting the Bachelors (BA) cycle (n=21; 70.0%) against a minority of Masters' (MA) students (n=9; 30.0%). Students came mostly from Eastern European countries, with a range of 6 different nationalities. These nationalities were: Austrian (n=1; 3.3%); Bielorussian (n=3; 10.0%); Lithuanian (n=2; 6.7%); Polish (n=17; 56.7%); Serbian (n=6; 20.0%) and Russian (n=1; 3.3%).

3.2.2. Scale characterization

Despite the scope of the questionnaire going beyond the assessment of the respondents’ stage of IC development, the centrepiece of this discussion is the group of difficulties arising from the adaptation of AIC and whether they derive exclusively from this process. Therefore, the instrument characterization and data analysis provided concern only the scale which assesses the respondents’ stage of IC development. This scale maintains the self-report structure of the AIC, including, however, 4 dimensions (instead of 5) and 21 items (instead of the 18 items for the 4 corresponding dimensions of AIC): attitudes (5 items); critical awareness (4 items); knowledge (7 items); skills (5 items). The adapted developmental level was the first – *Educational Traveller*.

Given the possible influence of the simple arrangement of response options on respondents’ ability to discriminate meaningfully (DeVellis, 2011), this characterization includes an example of the arrangement of the scale and response options in our questionnaire (Figure 1), where all dimensions were organized in a single table-list. Due to space constraints, the example in figure 1 represents only 1 of the 4 dimensions. The original measurement format, layout and colours are respected in this figure.

		0	1	2	3	4	5
		None	Very Low	Low	Neither high or low	High	Very high
Awareness	I am aware of differences across languages and cultures						
	I am aware of my negative reactions to these differences (fear, disgust superiority)						
	I am aware of how a specific context alters my interactions with others						
	I am aware of how I am viewed by members of the host culture						

Figure 1- Pilot questionnaire layout

3.2.3. Data analysis

The statistical software used to analyse the data was IBM SPSS (V.19). The first approach was a descriptive analysis of the dimensions’ scores, which were computed by considering the mean of the respondents’ answers to all items integrating each of the four IC dimensions. Furthermore, the reliability of the scale was assessed by computing Cronbach’s alpha, which is the most common reliability index for instruments using ordinal graduation-scales (Cronbach, 1984; DeVellis, 2011). The quality of the items was assessed through the inter-item correlations, which allowed us to ascertain which items should be weeded out and which should be retained.

Table 2 – Mean (M), Standard deviation (SD) and Cronbach’s alpha (α) for each dimension in the pilot questionnaire

Variables	N of valid answers per(*) item	M	SD	α
Attitudes	30*5	4.09	.55	.73
Critical awareness	30*4	3.07	.88	.59
Knowledge	30*7	3.52	.64	.72
Skills	29*5	3.53	.69	.68

Looking at the scores, it is noticeable the considerably high means per dimension, the “attitudes” variable being a striking example. As for the internal consistency of reliability indexes, the critical awareness dimension required cautious attention, even if the value was close to the .60 lower reliability bound suggested by DeVellis (2011).

A detailed examination of the scores, along with the think aloud reflections of a subgroup of respondents, led us to assume that one of the primary sources of error behind the high scores was the SDB associated with the scale, and so the likely non-equivalence between the IC construct underlying the AIC and our instrument. Thus, we hypothesised that this response style in which respondents portrayed themselves too positively (Johnson & Van de Vijver, 2003; Van de Vijver & Leung, 1997) was biasing the results.

Having identified a key measurement bias it was crucial to map the sources and unfolding nuisance factors, which were still undefined. Accordingly, we had to go through the data collection instrument again checking all the sources affecting measurement and outline a tailored step-by-step strategic plan.

3.3. Measurement errors and strategic plan

Tackling the liabilities of the pilot-study questionnaire was a time-consuming process with no straightforward answers. It led to sketching a four-fold strategic plan, since we unfolded the SDB into **construct bias** and **item bias**, later adding on two other sources of error which we labelled as “**meaningful discrimination**” and “**measurement format**”. Given that the nature of the four sources of bias determined the sort of solutions found, the type of biases will be discussed first and only then the solutions.

The recognition that our first ideas on how to measure IC were vaguer than we thought led us to realize that assessing the IC construct at general level was not the most appropriate choice, as it did not match the level of specificity of our research question. After all, the research setting was the Portuguese CE exchange context, *so why have a cultural-free construct?* Simply put, one of the sources of the bias was induced by the construct (**construct bias**), since the decentred approach (Werner & Campbell, 1970) of the AIC followed a culturally diverse perspective which did not offer a specific reference point respondents could relate to. This led us to question not only the cultural scope of our pool of items but also whether this pool offered an adequate sampling which covered the four dimensions of IC. In the view of the results, it did not indeed, as the 4 items added initially in the pilot questionnaire had undesirable similarity to other items and questionable relevance (De Vellis, 2011).

In addition to the incomplete construct coverage, the item formulation seemed to constitute a second source of bias (**item bias**), on account of the lack of equivalence between the still native-like level of English used and the non-native proficiency level of the respondents, as well as the frequent use of adjective forms.

Though there were no doubts about the two sources of bias reported, the physical arrangement of the question could also be jeopardizing its validity, because the disposition of both the scale and response options could be hindering the respondents’ ability to discriminate meaningfully. Moreover, the Likert scale seemed to be counter-productive, as respondents could be using the neutral midpoint (3= neither high nor low) as a “safety net”. Hence the addition of two other sources of error as summarized below.

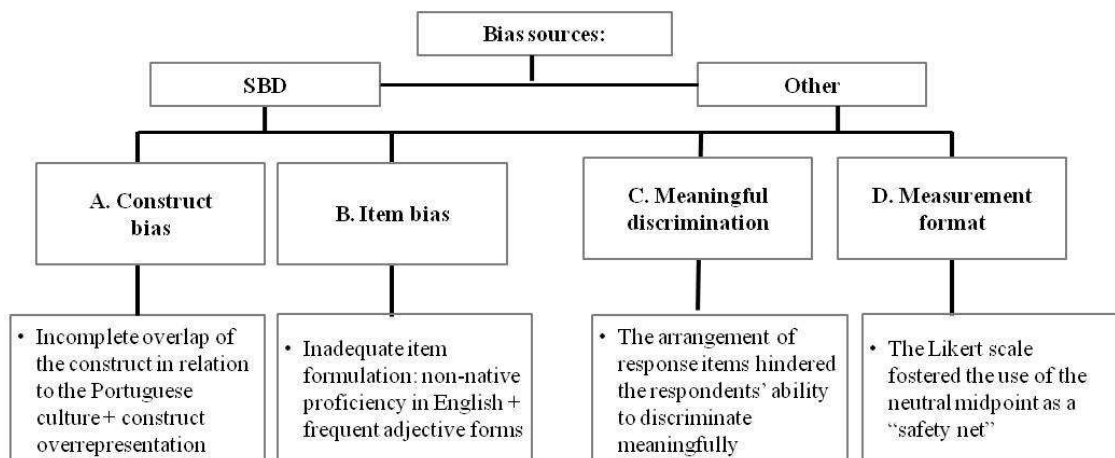


Figure 2 - Measurement errors

Upon the identification of the sources of bias and associated nuisance factors, we sought for the appropriate solutions. The nature of these problems frame the solutions found. These being:

- A. Construct bias:** adaptation of the scope of the IC scale to the Portuguese culture and item reduction in accordance to the number of items in the AIC, given that the original items covered the specificities of IC dimensions, but not in relation to the Portuguese culture. Furthermore, examples would be provided whenever the item could raise difficulties in interpretation;
- B. Item bias:** Lexical and syntactical reformulation of the items, in accordance with the respondents’ level of English (non-native proficiency) and their unfamiliarity with the intercultural jargon, which was eliminated. Verb forms were used instead of adjectives so as to enhance the focus of measurement on respondents’ performance and to lessen the socially desired attributes;
- C. Meaningful discrimination:** re-arrangement of item disposition by shadowing items alternately and providing the continuum of 6 levels of competence on each of the rows. Note that before the respondents had to tick one of the 6 degrees of competence, reading down the top row and ticking the suitable option across each of the items. Additionally, we also provided a visual representation of the scale, so as to highlight the underlying ascending continuum;
- D. Measurement format:** adoption of the original AIC scale, since the Likert scale seemed to create a neutral midpoint as a “safety net” when respondents did not want to precise their stage of IC.

3.4. The intervention instrument: data analysis

3.4.1. Participants

The final instrument was administered in November 2012 to 19 incoming CE students at UA in 2011/12. As this instrument precedes the intercultural training which took place in the PFL classroom, we were only able to administer it to 19 out of 25 CE movers. Eleven (11) of the participants were women (57.9%) and 8 were men (42.1%), with ages ranging between 20 and 25; average age 22.8 (SD= 1.4). As for the cycle of studies, the majority of students were doing the MA cycle (n=12; 63.2%) against a minority of BA students (n=7; 36.8%). Students came mostly from Eastern European countries once more, the nationality range comprising 5 different nationalities. These nationalities were: Austrian (n=1; 3.3%); Finnish (n=3; 10.0%); Lithuanian (n=2; 10.5%); Latvian (n=2; 10.5%) and Polish (n=11; 57.9%).

3.4.2. Characterization

The final scale comprises 19 items, 1 extra item in the awareness dimension in relation to both the pilot questionnaire and the AIC, as the first item of the awareness cluster was double-barrelled in that it conveyed two different (even if complementary) ideas. As for the measurement response format, we maintained the one in AIC. As in section 3.2.2, an example of the (re-arranged) layout of the intervention questionnaire is provided.

	No 0								5	Very high motivation
Awareness										
While in Portugal I became aware of...										
differences between my own language and the Portuguese language	↑	↓	↑	↓	↑	↓	↑	↓	↑	↓
differences between my own culture and the Portuguese culture	↑	↓	↑	↓	↑	↓	↑	↓	↑	↓
my stereotypes to these differences	↑	↓	↑	↓	↑	↓	↑	↓	↑	↓
how I behave differently according to the Portuguese culture	↑	↓	↑	↓	↑	↓	↑	↓	↑	↓
how Portuguese people see me (their opinion about me)	↑	↓	↑	↓	↑	↓	↑	↓	↑	↓

Figure 3- Intervention questionnaire layout

3.4.3. Comparative analysis and discussion

Ascertaining the effectiveness of the four-fold strategic plan requires a comparative statistical analysis of the data yielded both by the pilot (PQ) and intervention questionnaires (IQ).

Table 3 details the results of the descriptive statistics conducted.

Table 3- Descriptive scores of IC dimensions in PQ and IQ

Variables	PQ				IQ			
	N	Mdn	M	SD	N	Mdn	M	SD
Attitudes	30	3.90	4.09	.55	19	3.25	3.34	.55
Critical Awareness	30	3.13	3.07	.88	19	2.60	2.52	1.07
Knowledge	30	3.57	3.52	.64	19	2.80	2.86	.64
Skills	29	3.60	3.53	.69	19	3.20	3.17	.94

Considering the descriptive scores, two patterns emerge: both the means and medians decreased in all IC dimensions in the PQ. This decrease can be due to a decline in the SDB, stemming possibly from a more accurate and less overrated self-assessment. In addition, while the mode number of the observable variables (items) that compose the IC dimensions in the PQ was often 3, that is, the former neutral midpoint, in the IQ that did not happen.

The grounds for these results can be the effectiveness of the strategic plan and, thus, a decrease of the SDB in these dimensions, but also the different times of the questionnaire administration as well as the immersion lengths of the 2 groups of participants: 9 months for the Pilot group and 2 months and a half for the intervention group.

Given that the discussion of these results would not make sense without comparing the internal consistency level of the scale in the PQ and IQ, table 4 details the values of Cronbach's alphas in each of the four IC dimensions in both questionnaires.

Table 4- Cronbach's alpha (α) in PQ and IQ

Variables	Pre-test	
	PQ	IQ
Attitudes	.73	.00
Critical awareness	.59	.76
Knowledge	.72	.60
Skills	.68	.77

According to table 4, the “attitudes” and “knowledge” dimensions have the lowest alphas, which can be explained by the variability of the answers of the two groups of participants to the items within these 2 dimensions. In the particular case of the “attitudes” dimension, we had to pay careful attention to the unacceptable Cronbach's alpha yield, since the violation of reliability assumptions clearly demonstrated that something was wrong and we had to perceive the grounds for it. In fact, a .00 alpha is extremely distant from the acceptable lower bound for alpha suggested by well-known methodologists such as Nunnally (1978) and DeVellis (2011), who recommend .70 and .60 respectively.

In this study we follow the comfort ranges suggested by DeVellis (2011): “below .60 unacceptable; between .60 and .65, undesirable; between .65 and .70 minimally acceptable; between .70 and .80 respectable; between .80 and .90 very good; and much above .90, one should consider shortening the scale” (p. 109). Although these guidelines also suggest that something was not right in the attitudes dimension, our assumption was that the respondents' immersion was not long enough to yield consistent values and indicate an internal consistent dimension. After all, *could attitudes towards the Portuguese culture be completely activated in the beginning when the sojourning experience was starting?* This question makes even more sense because the two samples responded to the PQ and IQ in different points in time. Neither had been submitted to the intervention plan, but in the case of the pilot group the measuring point was at the end of their stay, when their attitudes towards the Portuguese culture were formed more consistently. Consequently, we formulated the hypothesis that the attitudes towards the host culture require a minimum immersion length. Testing this hypothesis would require re-administering the same scale in the post-test questionnaire. We took the risk and the result is **.63 for the same dimension** in the post-test.

Although it is imperative to look at the attitudinal items again, the results of the post-test do confirm the formulated hypothesis. This confirmation is enhanced by other studies such as Dwyer's (2004); Behrnd and Porzelt's (2012) that highlight the impact of the duration of a minimum of one-year stay in the development of IC.

4. Conclusions and implications

Although the answers to the two overarching questions raised in the introduction depend on the specificities of each study, sharing the decision-making process of any work committed to IC assessment is a compulsory step to bridge the gap between IC conceptualization and assessment. Disclosing ‘good’ and ‘bad’ decisions of the research activity is essential to increase the stock of knowledge on how to assess such a complex competence as IC.

The specific answers of this research are, then, the also specific solutions to the difficulties of the first phase of our assessment task. The underlying rationale can be, however, used as a “road map” for reaching a common destination: quality IC assessment, that is, assessment “undertaken for the purpose of improving student learning and development” (Palomba & Banta, 1999, p. 4). As for the “road map” it was for us, and still is, Deardorff’s (2009) recommendation of prioritising the “specific aspects of intercultural competence, based on the overall mission and purpose of the course, programme or organisation” (p. 481).

In the particular case of this study it was upon prioritizing that we were able to tailor the pre-test questionnaire to the specific purpose of our target intervention and reach consistent results for the “critical awareness” and “skills” dimensions. This process was aided not only by the depicted strategic-plan but also by involving the respondents in it, specifically through the think aloud method stated in section 3.2 and a question in the pre-test which determined the respondents’ thematic areas of interest for the intervention classes. The result was a series of topics striving for a more contextual and detailed understanding of the Portuguese culture.

Despite the different sizes of the two groups of participants and the conditions of the questionnaire administration, the preliminary data seems to indicate that culture-specific instruments can help precise the nomological networks underlying an instrument, as a more specific construct can lead to a more accurate interpretation. Also, even if the four-fold strategic plan contributed to the quality of our instrument (specially in the “awareness” and “skills” dimension), we cannot claim to know what is the impact of each strategy, on account of the simultaneous testing of the four strategies.

It is also important to note that though the centrepiece of this discussion was the adapted intercultural instrument, the wider PhD assessment plan does not rely solely on the pre and post-test questionnaires, but on a multi-method and multi-perspective approach. Thus, despite the quasi-experimental nature of the intervention plan, meaning the initial and final measurement points were of crucial importance, the ongoing assessment process was not disregarded. Hence a combination of multiple assessment types, formats and strategies, so as to obtain varied indicators of the students’ progress and attainment of the intervention learning objectives (Fantini, 2009). We refer in specific to: follow-up focus groups, video-recorded intervention classes, self-report essays and group works from the students, performative in-class tasks, along with the researcher’s logbook and the field notes from the PFL teacher.

.In summing up, even if the subjective nature of IC does not allow to completely eliminate the SDB in intercultural instruments (especially in self-report paper and pencil questionnaires), we believe that the IQ is, now, less susceptible to socially desirable response tendencies. Thus, we considered important to share the decision-making process of adapting and improving an IC assessment instrument which has been, for us, a useful benchmark in monitoring the specific aspects of IC that the PhD intervention aimed to facilitate.

In essence, there will always be subjectivity in assessing IC (Lussier & et al., 2007), yet scholarly effort in devising coherent assessment frameworks as well as sound external assessment instruments cannot be ignored. The risk is as simple as reinventing a wheel that could be much more sophisticated already.

References

- Behrnd, V. & Porzelt, S. (2012). Intercultural competence and training outcomes of students with experiences abroad. *International Journal of Intercultural relations*, 36(2), 213-223.
- Bhawuk, D.P.S. (1998). The role of culture-theory in cross-cultural training: A multimethod study of culture-specific, culture-general, and culture theory-based assimilators. *Journal of Cross-Cultural Psychology*, 29, 630-655.
- Bhawuk, D.P.S. (2001). Evolution of culture assimilators. Toward theory-based assimilators. *International Journal of Intercultural Relations*, 25, 141-163.
- Brown, C. & Knight, K. (1999). Introduction to self-awareness inventories. In S. M. Fowler & M. G. Mumford (Eds.), *Intercultural sourcebook: Cross-cultural training methods* (Vol.2, pp.19-30). Yarmouth: Intercultural Press.
- Cronbach, L. J. (1984). *Essentials of psychological testing*. New York: Harper & Row.
- Deardorff, D.K. (2004). The identification and assessment of intercultural competence as a student outcome of internationalization at universities of higher education in the United States. (Unpublished doctoral dissertation, North Carolina State University, 2004).
- Deardorff, D. K. (2005). A matter of logic. *International Educator*. 14(3), 26-31.
- Deardorff, D. K. (2006). Identification and assessment of intercultural competence as a student outcome of internationalization. *Journal of Studies in International Education*, 10(3), 241-266.
- Deardorff, D. K. (2009). Implementing intercultural competence assessment. In D.K. Deardorff (Eds.), *The sage handbook of intercultural competence* (pp. 456-476). Thousand Oaks: Sage.
- DeVellis, R.F. (2011). *Scale development: Theory and applications* (3rd edition). Thousand Oaks: Sage.
- Dwyer, M. M. (2004). More is better: The impact of study abroad program duration. *Frontiers: The Interdisciplinary Journal of Study Abroad*, 10, 151-164.
- Fantini, A. (2006). Exploring and assessing intercultural competence. Brattleboro, VT: Federation of the Experiment in International Living. Retrieved from www.worldlearning.org or www.experiment.org.
- Fantini, A. (2009). Assessing intercultural competence. Issues and tools. In D.K. Deardorff (Eds.), *The sage handbook of intercultural competence* (pp. 456-476). Thousand Oaks: Sage.
- Fowler, S. M. & Mumford, M. G. (1999). *Intercultural sourcebook: Cross-cultural training methods* (Vol.2). Yarmouth: Intercultural Press.
- Johnson, T. P. & Van de Vijver, F. (2003). Social Desirability in cross-cultural research. In J.A. Harkness, F. J. R. Van de Vijver & P. Ph. Mohler (Eds.), *Cross-cultural survey methods* (pp. 195-204). New Jersey: Willey.
- Lussier, D., Golubina, K., Ivanus, D., Kostova, S. C., Retamar, G.M., Skopinskaja, L. & Wiesinger, S. (2007). In I. Lazár, M. Huber-Kriegler, D. Lussier, G. S. Matei & C. Peck (Eds.), *Developing and assessing intercultural communicative competence* (pp. 23-37). Strasbourg: Council of Europe.
- Moreira, G. & Amorim, F. (2010). Hook Up! Enhancing the mobility of young Europeans through ICT and language competence. In Pixel (Ed.), *Proceedings of 3rd International Conference ICT for Language Learning*, Florence, Italy, 11-12 November 2010. Retrieved from http://www.pixel-online.net/ICT4LL2010/common/download/Proceedings_pdf/ILL06-Moreira,Amorim.pdf
- Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.
- Paige, M. (2004). Instrumentation in intercultural training. In D. Landis, J. M. Bennett & M. J. Bennett (Eds.), *Handbook of intercultural training* (pp. 85-128) (3rd edition). Thousand Oaks: Sage.
- Palomba & Banta. (1999). *Assessment essentials: Planning, implementing and improving assessment in higher education*. San Francisco: Jossey-Bass Publishers.
- Papatsiba, V. (2006). Making higher education more European through student mobility? Revisiting EU initiatives in the context of the Bologna process. *Comparative Education*, 42(1), 93-111.
- Szarka, J. (2003). Student mobility in the EU. *European Integration Studies*, 2 (2), 123-138.
- Vande Berg & Paige (2009). The evolution of intercultural competence in U.S. study abroad. In D.K. Deardorff (Eds.), *The sage handbook of intercultural competence* (419-437). Thousand Oaks: Sage.
- Van de Vijver, F. & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Thousand Oaks: Sage.
- Van de Vijver, F. & Leung, K. (2009). Methodological issues in researching intercultural competence. In D.K. Deardorff (Eds.), *The sage handbook of intercultural competence* (pp. 405-418). Thousand Oaks: Sage..
- Wächter, B. (2010). Mobility and internationalization in the European Higher Education Area. In M. Kelo (Eds.), *Beyond 2010. Priorities and challenges for higher education in the next decade* (pp. 13-42). Bonn: Lemmens.
- Werner, O. & Campbell, D.T. (1970). Translating, working through interpreters, and the problem of decentering. In R. Naroll & R. Cohen (Eds.), *A handbook of cultural anthropology* (pp. 398 -419). New York: American Museum of Natural History.