MA. Jose Fabian Elizondo-Gonzalez[1]
Universidad de Costa Rica

Original scientific paper
UDC: 37.043.2
DOI: 10.5937/IstrPed2402377E

==========

ANALYZING IMMIGRANT AND NON-IMMIGRANT BELONGING EXPERIENCES THROUGH IRT IN LARGE-SCALE ASSESSMENTS: INSIGHTS FROM COSTA RICA

Abstract: Studies across countries part of the Organization for Economic Cooperation and Development (OECD) suggest that learners' sense of school belonging is often influenced by their place of birth. However, large-scale assessment studies rarely explore whether differences in belonging scores between non-immigrant and immigrant learners are due to test bias. This study fills that gap by examining belonging scores in Costa Rican high schools using the Programme for International Student Assessment (PISA) 2022 data, with a focus on test fairness. Utilizing Multiple-group Differential Item Functioning (DIF) analyses and Item Response Theory (IRT) modeling, results show that all items are DIF-free, confirming no test bias. An Independent Samples t-test reveals no significant differences in belonging scores between immigrant and non-immigrant learners, which is a positive finding. It suggests that Costa Rican educational environments foster a shared sense of belonging, regardless of learners' place of birth. The most discriminating items, identified through IRT modeling, relate to performative or participatory aspects of school belonging. This study highlights the importance of incorporating IRT modeling and fairness protocols in large-scale assessments. By confirming that sense of belonging is not impacted by place of birth, stakeholders can confidently make decisions that further support inclusive educational practices in Costa Rican classrooms, knowing the PISA sense of belonging index provides unbiased, reliable scores.

Keywords: belonging, Costa Rica, Differential Item Functioning, immigrant learners, test bias

Introduction

As reported by the Organization for Economic Cooperation and Development (OECD) (2023d), "the share of immigrant students has increased in Costa Rica to 12% in 2022 (6% in 2012). In 2022, 4% of 15-year-old students were first-generation immigrants" (par. 7). This considerable increase in the number of immigrant learners in Costa Rican classrooms cannot go unnoticed. These learners might encounter multiple hurdles trying to incorporate into a new culture, such as socializing, integrating, and succeeding academically in a foreign educational environment. Hence, one of the elements in which research can help determine if immigrant learners have a comparable thriving environment to their Costa Rican counterparts is by comparing both groups' sense of belonging at school. Using the Student Performance data from the PISA 2022 data, the researcher will aim at determining whether there is a Differential Test Functioning (DTF) for non-immigrant and immigrant learners, if there is a statistical mean difference across groups, and at analyzing item discrimination and item location parameters with these two sample populations.

[1] josefabian.elizondo@ucr.ac.cr

1. Literature review

Multiple authors have defined sense of belonging through a myriad of dimensions, behaviors, skills, attitudes, and scales. One of the most recent definitions is proposed by Allen, Kern, Rozek, McInerney, & Slavich (2021), who posit that four components make up sense of belonging: competencies for belonging, opportunities to belong, motivations to belong, and perceptions of belonging, all influenced by a person's environment and context (p. 92). Similarly, Mahar, Cobigo, and Stuart (2013) analyzed over 8000 articles on sense of belonging and created their own conceptualizations of its elements based on the common themes found in their search to provide a "transdisciplinary, multidimensional understanding of the sense of belonging" (p. 1030). These elements are subjectivity, groundedness, reciprocity, dynamism, and self-determination. In short, they relate to an individual's perception of feeling valued, belonging to something, sense of connectedness, relation to their physical and social environments, and their choice to control to whom and what to belong to (Mahar et al., 2013: 1030-1031). Therefore, based on these definitions not only do individuals' actions play a role in developing a sense of belonging, but also their surrounding environments and contexts.

More specifically within the educational context, Goodenow (1993) has defined school belonging as "the extent to which students feel personally accepted, respected, included and supported by others in the school social environment" (p. 80). Based on this concept, Goodenow (1993) has created the Psychological Sense of School Membership Scale (PSSM), one of the most used scales today. Using it as a model, others have adapted and expanded their own definitions and instruments to operationalize school belonging; for example, Elizondo (2023) posits that school belonging could potentially be made of internal and external (performative) manifestations, while St-Amand, Bowen, & Wan Jung Lin, (2017) propose in their scale that school belonging is composed of and measured in multiple dimensions such as the emotional, social, participatory, and adaptive ones. Another scale that has taken the PSSM as a model is the PISA sense of belonging index. This six-item scale has its foundation on the theories proposed by Maslow (1943), Goodenow (1993), Baumeister and Leary (1995), and Ma (2003), where belonging represents a "need to form and maintain at least a minimum number of interpersonal relationships based on trust, acceptance, love and support" (OECD, 2019, p. 130). This index has been administered internationally for over two decades to hundreds of thousands of 15-year-old learners, mostly from OECD member countries.

Investigating sense of school belonging in immigrant populations is paramount since "students with an immigrant background in many education systems are at an increased risk of academic underperformance, a low sense of belonging to their school community and low life satisfaction" (Cerna, Brussino, & Mezzanotte, 2021: 64). Numerous studies have shed light on group differences in terms of sense of beloging and immigration status. In Europe, Rodriguez et al. (2020) found that Spanish non-immigrant learners scored statistically significantly higher in feelings of beloging at school over their first- and second-generation immigrant counterparts (p. 96). Likewise, He and Fisher (2020) reported that using the PISA data from 2015 there was a statistically significantly higher feeling of belonging of non-immigrant over migrant learners in Spain, Italy, and Germany (p. 6). Cerna, Brussino, and Mezzanotte (2021) also found that immigrant leaners were less likely to report a higher sense of belonging over non-immigrants in countries such as Germany, Luxerbourg, Portugal, and Sweden (p. 28). This analysis was done with items ST034Q01TA (I feel like an outsider at school) and ST034Q03TA (I feel like I belong at school), based on the PISA 2018 results. A more detailed account is provided by Hogberg, Petersen, Strandh, and Johansson (2021) after analyzing the PISA results in Sweden from 2000-2018, where they uncovered that "2012 onwards, foreign-born students had clearly and significantly lower belonging on average" (p. 973). Interestingly, Cerna, Brussino, and Mezzanotte (2021) reported that immigrant learners in other latitudes had experienced the opposite: in Australia, New Zealand, and the United Kingdom, immigrant learners "were more likely to report a greater sense of belonging at school than native students" (p. 28). In America, the results are mixed. Cerna, Brussino, and Mezzanotte (2021) found no statistically

significant differences in sense of belonging among non-immigrant and immigrant learners in Costa Rica, Chile, and the US, but they did find a statistically significant difference among these groups in Canada (p. 30).

Fortunately, based on the PISA 2018 results, "across OECD countries, the majority of students reported that they feel socially connected at school" (OECD, 2019, p.130). In addition, the PISA 2022 results also report that in these countries, learners who were not exposed to bullying at school and who enjoyed more family support demonstrated a stronger sense of belonging at school (OECD, 2023c, n.p.). Since COVID-19, however, these results may have changed. Based on the PISA 2022 report, it is stated that "students in systems that spared more students from longer closures […] reported a greater sense of belonging at school" (OECD, 2023c, n.p.), and countries such as Costa Rica had full closures "for an average of 175 days between 1 January 2020 and 20 May 2021" (OECD iLibrary, 2024, n.p.). Consequently, the effects of these closures are to be examined and furthered studied, considering these recent life-changing, worldwide events, which, as discussed, could have introduced more sources of increased risks for a lower sense of school belonging in immigrant learners, even where there had not been any differences across groups in the past.

Even though there are a myriad of articles that compare non-immigrant and immigrant learners using PISA's belonging index, they do not tend to address if one of the reasons for score differences is due to test bias. Test bias is relevant in this context since it could lead "to potentially misleading results with regards to group differences" (Langer et al., 2008: 3). One of the few studies that does employ test bias analyses when studying the impact of immigrant backgrounds using the PISA belonging index is that of Roberson (2020), who affirms that "measurement theorists should also consider how migration background should be taken into account not just in terms of measurement invariance, but the way in which one's migration background actually (re) shapes the constructs at hand" (p. 73). In his study, he confirms the presence of DIF (differential item functioning), on three of the six items, based on students' immigrant backgrounds when analyzing a sample of 83,300 students from 14 different countries, a conclusion that lead him to report that "it is *inappropriate* to treat all peoples of different immigration backgrounds as equivalent on the PISA measure of belonging" (Roberson, 2020: 70).

There are multiple methods to analyze test and item bias. Differential Test Functioning (DTF) is observed when many items on a test favor or hinder a "characteristic exhibited by group members of a test-taking population" (Runnels, 2013: 1). To illustrate, an item that has bias could make a person from an immigrant background, but with the same ability level as a learner with a non-immigrant background, score lower on a test just because of that person's group membership. Item bias, or DIF, is understood as "when the parameter estimates in the first group are not the same as the estimates in the second, [and then] the item is considered to be functioning differentially in the two groups" (Kim, Cohen, & Park, 1995: 262). Nugent (2017) states that "DIF can accumulate across items on a scale to create DTF", and that in the absence of DIF "the test characteristic curves […] will be the same" (pp. 305-317). For that reason, one method to decide if DIF items cause DTF is by analyzing those DIF items' effect sizes, which can be done by determining if there is a statistical difference among the areas under the combined Test Characteristic Curves (TCCs), resulting from the Multiple-group DIF analyses.

Multiple-group DIF analysis can help identify DIF when comparing two groups' parameters jointly (Bock & Zimowski, 1997). Multiple-group DIF analysis can be approached in two ways: the constrained baseline model and the free baseline model (Wang, 2004 as cited in Chun, Stark, Kim, & Chernyshenko, 2016). In the constrained baseline model items are estimated with an equality constraint, except for the studied item. If there is a significant difference between two groups' item parameters in the studied item, one can determine this item has DIF. In the free baseline, all items, except for anchor items, are freely estimated. Just like with the constrained baseline model, if there is a significant difference between two groups' item parameters in the studied item, one can

determine this item has DIF. Research has shown that combining these approaches reduces type I error, or the possibility of detecting an item as DIF when it is not, since the free baseline approach can adjust for the type I errors obtained in the constrained baseline model (Stark et al., 2006), while it also can make use of the DIF-free items identified in the constrained baseline model as anchor items. Consequently, researchers can start their analysis with the constrained baseline model to identify DIF-free items and then use those as anchor items for the free baseline model, where they will confirm statistically if the DIF items found in the constrained baseline mode have in fact DIF or if they were a product of a type I error.

2. Research questions

This research paper attempts to answer the following questions:

1. Is there a Differential Test Functioning (DTF) for non-immigrant and immigrant learners in the PISA sense of belonging items?
    a. Is there any Differential Item Functioning (DIF) in the sample? If so, how many items show DIF?
    b. What do Test Characteristic Curves, Test Information Function plots, and Standard Error plots indicate?
2. Are the mean scores of immigrant learners statistically different from the mean scores of non-immigrant learners?
3. Are there differences in terms of item discrimination and item location in the PISA sense of belonging items for non-immigrant and immigrant learners?

3. Methods[2]

PISA 2022 data, obtained from the official OECD database (OECD, 2024), was subset to containing only data reported by learners in Costa Rica on the following variables: Country code (CNT), learners' place of birth (ST019AQ01T), sex assigned at birth (ST004D01T), sense of belonging items (ST034Q01TA, ST034Q02TA, ST034Q03TA, ST034Q04TA, ST034Q05TA, ST034Q06TA). 6113 observations on 9 variables were obtained. For reference, the researcher used the "PISA 2022 Compendia on Questionnaire items" and the "Student questionnaire" for the coding of the variables, where non-immigrants were coded as 1 in the original dataset, while immigrants as 2 and 3. Females were coded as 1 and males as 2 (OECD, 2023a, n.p.). Finally, the six items on sense of belonging were formatted in a four-point Likert scale, ranging from 1 to 4 (Strongly agree, Agree, Disagree, and Strongly Disagree). Table 1 shows the items included in this PISA 2022 administration.

Table 1. *PISA 2022 Sense of Belonging Items*

Item code	Item given to learners
ST034Q01TA	I feel like an outsider (or left out of things) at school.
ST034Q02TA	I make friends easily at school. *
ST034Q03TA	I feel like I belong at school. *
ST034Q04TA	I feel awkward and out of place in my school.
ST034Q05TA	Other students seem to like me. *
ST034Q06TA	I feel lonely at school.

Note. *These items had to be reverse coded before the initial analysis so that higher values reflected higher school belonging.

To analyze the data, a Generalized Partial Credit Model (GPCM) was employed as it is PISA's choice for IRT (Item Response Theory) model analysis "in the case of items with more than two categories" (OECD, 2019, p. 212). However, after conducting deviance-based model fit statistics and

[2] This study was a secondary analysis of de-identified data that is publicly available. Consequently, it did not undergo an ethical review by an Institutional Review Board (IRB).

a Likelihood-ratio test, it may seem that for this sample the Graded Response Model (GRM) could have been a more suitable fit, since out the tested models (PCM, GPCM, and GRM), the GRM had the smallest values in AIC, BIC, and SABIC of the three. Table 2 summarizes the multiple values for the deviance-based model fit statistics and the Likelihood-ratio test. The researcher still employed the GPCM as to be consistent with the IRT methodology employed by PISA and other studies that analyze these datasets.

Table 2. *Deviance-based model fit statistics and the Likelihood-ratio test results*

Model/results	AIC	BIC	SABIC	logLik	χ^2	df	p
PCM	54463.45	54588.36	54527.99	-27212.73	-	-	-
GPCM	54282.86	54440.64	54364.38	-27117.43	190.597	5	0
GRM	53635.43	53793.21	53716.95	-26793.71	647.428	0	NaN

To quantify if any items presented item bias, Differential Item Functioning (DIF) was analyzed through Multiple Group IRT DIF analysis (Bock & Zimowski, 1997; Reckase, 2009) using the *mirt* package in R. To account for group differences, a two-tailed Independent Samples t-test was computed and analyzed the .05 alpha level, where the mean scores on sense of belonging of non-immigrant learners (coded 1) would be contrasted against the mean scores of immigrant learners (coded as 2 and 3). Finally, step parameters from the GPCM model had an alternative parameterization, where their slope-intercept parameters were converted into IRT location-threshold parameters to report item discrimination and item location in both groups.

4. Results

4.1. Descriptive statistics

For this dataset, there were 6113 observations, but 176 learners did not report their place of birth, and 643 did not report data for any of the six items. These rows were deleted list-wise since they could not provide information for categorizing the learners or for data analysis. The new total was 5294 observations, from which 5016 were non-immigrant and 278 were immigrant leaners; 2662 were females and 2632 males. Table 3 shows the counts of the intersection of both non-immigrant and immigrant, female and male learners in the dataset.

Table 3. *Number of learners based on place of birth and sex assigned at birth*

Counts based on place of birth		Counts based Gender distribution	
Non-immigrant	5016	Female	2662
Immigrant	278	Male	2632
Totals	5294		

Table 4 provides the descriptive statistics (average score) for the non-immigrant learners, and Table 5 provides a summary of the descriptive statistics for each of the items taken by these participants.

Table 4. *Descriptive statistics for non-immigrant learners*

n	*M*	*SD*	*Mdn*	*Min*	*Max*	*Range*	*Skew.*	*Kurt.*	*SE*
5016	2.99	.67	3	1	4	3	-.71	.49	.01

Table 5. *Descriptive statistics for sense of belonging items for non-immigrant learners*

Item	*n*	*M*	*SD*	*Mdn*	*Skew.*	*Kurt.*	*SE*	Number of NAs
ST034Q01TA	4054	3.04	.93	3	-.77	-.22	.01	962
ST034Q02TA	4164	2.06	.89	2	.60	-.32	.01	852
ST034Q03TA	4056	1.99	.82	2	.74	.28	.01	960
ST034Q04TA	4091	2.98	.90	3	-.71	-.17	.01	925

ST034Q05TA	4145	2.05	.79	2	.68	.38	.01	871
ST034Q06TA	4083	3.04	.91	3	-.78	-.12	.01	933

Table 6 provides the descriptive statistics for the immigrant learners, and Table 7 provides a summary of the descriptive statistics for each of the items taken by these participants.

Table 6. *Descriptive statistics for immigrant learners*

n	*M*	*SD*	*Mdn*	*Min*	*Max*	*Range*	*Skew.*	*Kurt.*	*SE*
278	2.93	.68	3	1	4	3	-.64	.2	.04

Table 7. *Descriptive statistics for sense of belonging items for immigrant learners*

Item	*n*	*M*	*SD*	*Mdn*	*Skew.*	*Kurt.*	*SE*	Number of NAs
ST034Q01TA	223	2.99	.93	3	-.72	-.29	.06	55
ST034Q02TA	224	2.12	.93	2	.55	-.51	.06	54
ST034Q03TA	223	2.00	.84	2	.73	.14	.06	55
ST034Q04TA	227	2.95	.94	3	-.73	-.29	.06	51
ST034Q05TA	227	2.15	.77	2	.64	.33	.06	51
ST034Q06TA	221	2.96	.87	3	-.59	-.24	.06	57

4.2 Inferential Statistics

Research Question 1: Is there a Differential Test Functioning (DTF) for non-immigrant and immigrant learners in the PISA sense of belonging items?

After conducting a Multigroup IRT DIF analysis, with a constrained baseline model, the adjusted p-value from the Likelihood ratio test proves to be not significant for all items, which indicates no presence of items with DIF. Consequently, all items are considered DIF free. Table 8 summarizes the results from the DIF analysis using the constrained baseline model.

Table 8. *DIF detection from the constrained baseline model*

Items	Groups	AIC	SABIC	HQ	BIC	X2	df	p	adj_p
ST034Q01TA	1, 2	7.738	21.325	16.928	34.036	0.262	4	0.992	0.992
ST034Q02TA	1, 2	6.207	19.794	15.397	32.504	1.793	4	0.774	0.929
ST034Q03TA	1, 2	6.126	19.713	15.317	32.424	1.874	4	0.759	0.929
ST034Q04TA	1, 2	4.033	17.620	13.223	30.331	3.967	4	0.411	0.929
ST034Q05TA	1, 2	5.849	19.435	15.039	32.146	2.151	4	0.708	0.929
ST034Q06TA	1, 2	0.787	14.374	9.977	27.085	7.213	4	0.125	0.750

Figure 1, Figure 2, and Figure 3 display the combined Test Characteristic Curves (TCCs), Test Information Function (TIF) plots, and Standard Error (SE) plots to differentiate learners' sense of belonging based on their group membership. The darker lines correspond to the non-immigrant group (1) and the lighter lines to the immigrant group (2). These plots show that after confirming that there is no item bias, therefore no test bias, these two samples behave exactly alike when reporting their sense of belonging since the lines in the plots are identical and superposed.

Figure 1. *Combined TCCs*

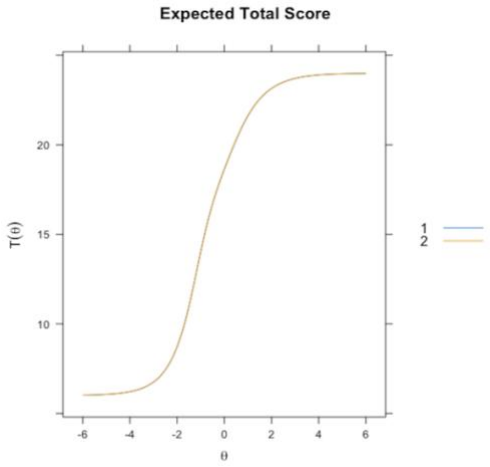


Figure 2. *Combined TIF plot*

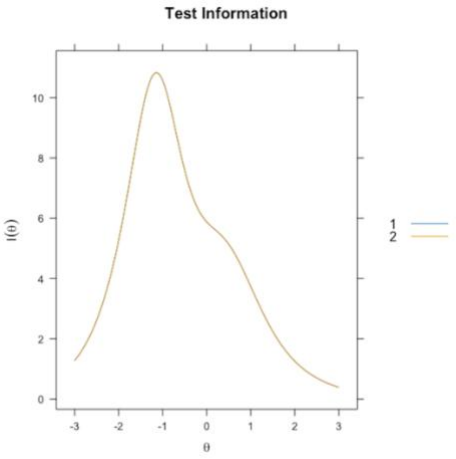
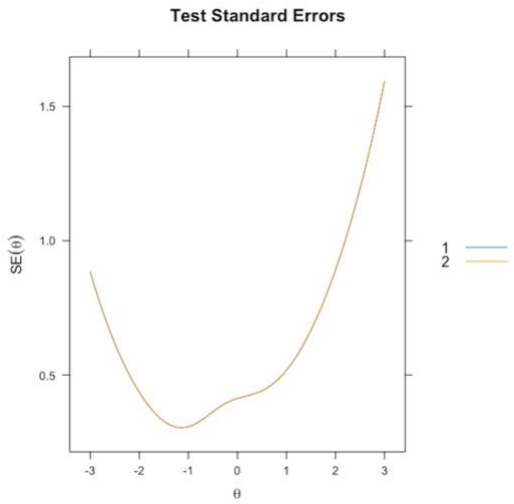


Figure 3. *Combined SE plot*



More specifically, Figure 1 shows that learners with a low sense of belonging[3] are expected to obtain fewer than 15 points as their total score, especially those at the -2-theta level. Figure 2 illustrates where the test provides the most information for people at different theta levels. In this figure both groups follow the same distribution, where the items provide the most information at the -1-theta level, providing less information starting at the 0-theta level. Finally, in Figure 3, the SE plot indicates a measurement of accuracy and reliability. For both groups, the test seems to be the most accurate and reliable at the -1 to 1 theta levels, which implies that the test is not the most accurate to measure sense of belonging for learners who are either located below the -1-theta level or above the 1 theta level, introducing even more inaccuracy at the higher theta levels.

Research Question 2: Are the mean scores of immigrant learners statistically different from the mean scores of non-immigrant learners?

[3] Sense of belonging (latent trait) will be referred to as "theta" in these analyses.

A Levene's test was conducted to assess the homogeneity of variance between the two groups based on the row totals of the specified items. The test revealed that the assumption of homogeneity of variance was not violated, as indicated by a non-significant result, F (1, 5935) = 2.2937, p= .13.

An Independent Samples t-test was conducted to compare the mean scores of non-immigrant and immigrant leaners' sense of belonging scores. There was no significant difference in scores between non-immigrant learners (M = 2.99, SD = 0.67) and immigrant learners (M = 2.93, SD = 0.68); t (307.88) = 1.51, p = .13, 95% CI [-0.02, 0.15]. This suggests that there is no evidence to support a difference in the mean scores between non-immigrant and immigrant leaners in this sample.

Research question 3: Are there differences in terms of discrimination and item location in the PISA sense of belonging items for non-immigrant and immigrant learners?

To analyze item discrimination (a) and item location (b), the step parameters obtained using the Multiple group IRT DIF analysis with a constrained baseline model and GPCM were converted to location-threshold parameters. Since all items are DIF free, the parameters are the same for both groups. Table 9 shows a comparison of the GPCM step parameters and their parameterization into location-threshold parameters.

Table 9. *GPCM step parameters and location-threshold parameters for both groups*

	GPCM step parameters				GPCM location-threshold parameters				
Item/parameter	a	b1	b2	b3	a	b	d1	d2	d3
ST034Q01TA	2.13	-1.280	-1.014	.365	2.13	.30	-.30	-0.17	.47
ST034Q02TA	1.15	-1.512	-1.159	.774	1.15	.55	-.77	-0.46	1.22
ST034Q03TA	1.31	-1.578	-1.444	.832	1.31	.56	-.65	-0.55	1.19
ST034Q04TA	1.88	-1.287	-1.046	.598	1.88	.31	-.38	-0.25	.63
ST034Q05TA	1.05	-1.781	-1.603	1.288	1.05	.67	-1.03	-0.86	1.90
ST034Q06TA	2.13	-1.321	-1.031	.433	2.13	.30	-.32	-0.18	.50

Item ST034Q05TA— Other students seem to like me—had the highest location for both groups, and it behaved as the least discriminating item. On the other hand, items ST034Q01TA— I feel like an outsider (or left out of things) at school— and ST034Q06TA—I feel lonely at school— showed to be the most discriminating and lowest location items in the two groups.

5. Discussion

The fact that all items are DIF free confirms that there is no test bias, or Test Differential Functioning (DTF) between non-immigrant and immigrant learners in Costa Rica. Since there are no DIF items identified, it can be concluded that whatever difference found between groups is real, not induced by test bias. In this sample, the result from the Independent Samples t-test shows that there is not a statistically significant difference between the two groups. Hence, it can be inferred that school sense of belonging might not be impacted by learners' place of birth: both Costa Rican learners and those born in other countries are expected to develop and report a comparable sense of belonging in schools, which supports the findings about Costa Rica in Cerna, Brussino, and Mezzanotte (2021) from the PISA 2018 results. The finding from the Independent Samples t-test is graphically represented in the combined Test Characteristic Curves (TCC), Test Information Function (TIF) plots, and Standard Error (SE) plots, which show that at the same theta levels learners from both groups are expected to obtain the same total scores, while the test yields the most level of information and has the least error at the -1 theta level. It should be noted, though, that this set of items may not be the most suitable instrument to capture sense of belonging of participants with higher theta levels, which can be observed in the steep line in the SEM plot (Figure 3) above the 1 theta level.

Because there is no DIF, item discrimination (a) and item location (b) parameters amongst the two groups are identical. Item ST034Q05TA—Other learners seem to like me— proves to be the least discriminating, yet it is the item with highest location, which best measures learners with a lower sense of belonging. An explanation for this might be that for people with lower levels of school belonging in this sample what others think of a specific individual seems to be a proxy of not belonging to the school community, though the item itself provides the least contribution in discriminating learners based on their belonging levels. Conversely, items ST034Q01TA and ST034Q06TA have the lowest item location and are the most discriminating items across participants. What this means is that these two items best measure learners with a high level of school belonging while also best discriminate amongst learners from low to high levels of school belonging. Contrary to item ST034Q05TA, these two items relate to internal feelings and perceptions of individuals within educational communities. They both refer to individuals' feeling like outsiders or being lonely at school. These items reflect dimensions of school belonging where learners' successful integration and participation in school communities is hindered, limiting their engagement and forming of bonds with people in their groups, which can be seen as a compromised actualization of performative or social/participatory dimensions of sense of belonging (Elizondo, 2023; St-Amand et al., 2017).

6. Conclusions

The analysis revealed no significant Differential Test Functioning between non-immigrant and immigrant learners in Costa Rican schools, indicating that if there were any differences between the two groups, these would be genuine disparities in sense of belonging. However, the Independent Samples t-test showed no statistically significant differences between the groups, which can be visually inspected in the identical, superposed lines in the combined TCC, TIF, and SE plots. Item analysis demonstrated that items related to internal feelings of exclusion better measured learners with high levels of school belonging and also had higher discriminatory power than those concerning peer acceptance. In sum, this study provides insights into the intricate dynamics of school belonging among non-immigrant and immigrant learners in Costa Rican high schools. By employing rigorous analytical techniques and drawing on a robust dataset, the findings offer nuanced perspectives on the factors shaping learners' sense of belonging, laying the groundwork for future research and informed interventions aimed at promoting educational equity and social inclusion, especially in Latin American contexts.

Declarations of interest: none

Declaration of Generative AI and AI-assisted technologies in the writing process: During the preparation of this work, the author used ChatGPT3 to write, adapt, and troubleshoot R code as well as to improve the readability of the abstract and conclusion. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the content of the publication.

Funding: This work was supported by *Universidad de Costa Rica* through the sponsorship of the author's current enrollment in the Educational Psychology and Research PhD. program at the University of Kansas, USA.

References

Allen, K. A., Kern, M. L., Rozek, C. S., McInerney, D. M., & Slavich, G. M. (2021). Belonging: A review of conceptual issues, an integrative framework, and directions for future research. *Australian Journal of Psychology*, 73(1), 87-102. DOI:10.1080/00049530.2021.1883409

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, *117*(3), 497–529. https://doi.org/10.1037/0033-2909.117.3.497

Bock, R.D., Zimowski, M.F. (1997). Multiple Group IRT. In: van der Linden, W.J., Hambleton, R.K. (Eds.), *Handbook of Modern Item Response Theory* (pp. 433-448). New York, NY: Springer. https://doi.org/10.1007/978-1-4757-2691-6_25

Cerna, L., Brussino, O., & Mezzanotte, C. (2021). The resilience of students with an immigrant background: An update with PISA 2018. *OECD Education Working Papers*, *261*, 1-69. https://dx.doi.org/10.1787/e119e91a-en

Chun, S., Stark, S., Kim, E. S., & Chernyshenko, O. S. (2016). MIMIC methods for detecting DIF among multiple groups: Exploring a new sequential-free baseline procedure. *Applied Psychological Measurement*, *40*(7), 486-499. DOI:10.1177/0146621616659738

Elizondo, J.F. (2023). Creating a validity argument for the" KU-Sense of Belonging Test" for exchange graduate students. *Research in Pedagogy*, *13*(2). DOI:10.5937/IstrPed2302403G

Goodenow, C. (1993). The psychological sense of school membership among adolescents: Scale development and educational correlates. *Psychology in the Schools*, *30*(1), 79-90. https://psycnet.apa.org/doi/10.1002/1520-6807(199301)30:1%3C79::AID-PITS2310300113%3E3.0.CO;2-X

He, J., & Fischer, J. (2020). Differential associations of school practices with achievement and sense of belonging of immigrant and non-immigrant students. *Journal of Applied Developmental Psychology*, *66*, 1-12. https://doi.org/10.1016/j.appdev.2019.101089

Hogberg, B., Petersen, S., Strandh, M., & Johansson, K. (2021). Determinants of declining school belonging 2000–2018: The case of Sweden. Social Indicators Research, 157(2), 783-802. https://doi.org/10.1007/s11205-021-02662-2

Kim, S. H., Cohen, A. S., & Park, T. H. (1995). Detection of differential item functioning in multiple groups. *Journal of Educational Measurement*, *32*(3), 261-276. DOI:10.1111/jedm.12415

Langer, M. M., Hill, C. D., Thissen, D., Burwinkle, T. M., Varni, J. W., & DeWalt, D. A. (2008). Item response theory detected differential item functioning between healthy and ill children in quality-of-life measures. *Journal of clinical epidemiology*, *61*(3), 268-276. DOI:10.1016/j.jclinepi.2007.05.002

Ma, X. (2003). Sense of belonging to school: Can schools make a difference? *The Journal of Educational Research*, *96*(6), 340-349. DOI:10.1080/00220670309596617

Mahar, A. L., Cobigo, V., & Stuart, H. (2013). Conceptualizing belonging. *Disability and rehabilitation*, *35*(12), 1026-1032. DOI:10.3109/09638288.2012.717584

Maslow, A. H. (1943). A theory of human motivation. *Psychological Review*, *50*, 370-396. https://doi.org/10.1037/h0054346

Nugent, W. R. (2017). Understanding DIF and DTF: Description, methods, and implications for social work research. *Journal of the Society for Social Work and Research*, *8*(2), 305-334. DOI:10.1086/691525

Organisation for Economic Co-Operation and Development [OECD]. (2024). *PISA 2022 Database.* https://www.oecd.org/en/data/datasets/pisa-2022-database.html

Organisation for Economic Co-Operation and Development [OECD]. (2023a). *PISA 2022 Compendia- Questionnaire Items.* https://webfs.oecd.org/pisa2022/PISA2022_FinalRelease_Compendia_5Dec23.zip

Organisation for Economic Co-Operation and Development [OECD]. (2023b). *Education at a*

Glance: OECD Indicators. Retrieved from: https://www.oecd-ilibrary.org/sites/9cc07f93-en/index.html?itemId=/content/component/9cc07f93-en

Organisation for Economic Co-Operation and Development [OECD]. (2023c). *PISA 2022 Results (Volume II): Learning During – and From – Disruption*. https://doi.org/10.1787/a97db61c-en

Organisation for Economic Co-Operation and Development [OECD]. (2023d). *Education GPS*. https://gpseducation.oecd.org/CountryProfile?primaryCountry=CRI&treshold=10&topic=PI

Organisation for Economic Co-Operation and Development [OECD]. (2019). *PISA 2018 results (volume III): What school life means for students' lives*. https://doi.org/10.1787/acd78851-en

OECD iLibrary. (2024). *Costa Rica*. https://www.oecd-ilibrary.org/sites/9cc07f93-en/index.html?itemId=/content/component/9cc07f93-en

Reckase, M.D. (2009). *Multidimensional Item Response Theory*. New York: Springer-Verlag.

Roberson, N. D. (2020). *On the measurement of social belonging and its connection to migration background* (T). University of British Columbia. Retrieved from https://open.library.ubc.ca/collections/ubctheses/24/items/1.0389711

Rodriguez, S., Valle, A., Gironelli, L. M., Guerrero, E., Regueiro, B., & Estevez, I. (2020). Performance and well-being of native and immigrant students. Comparative analysis based on PISA 2018. *Journal of Adolescence*, *85*, 96-105. https://doi.org/10.1016/j.adolescence.2020.10.001

Runnels, J. (2013). Measuring differential item and test functioning across academic disciplines. *Language Testing in Asia*, *3*, 1-11. DOI:10.1186/2229-0443-3-9

Stark, S., Chernyshenko, O. S., & Drasgow, F. (2006). Detecting differential item functioning with confirmatory factor analysis and item response theory: toward a unified strategy. *Journal of Applied Psychology*, *91*(6), 1292.

St-Amand, J., Bowen, F., & Wan Jung Lin, T. (2017). Le sentiment d'appartenance à l'ecole : une analyse conceptuelle. *Canadian Journal of Education/Revue canadienne de l'education*, *40*(1), 1-32.

Biographical note:

Jose Fabian Elizondo Gonzalez is an EFL instructor and researcher who works at Universidad de Costa Rica. He holds two master's degrees, one in Education Administration and a second one in Teaching English as a Foreign Language. Currently, he is finishing his third year as a PhD student in the Educational Psychology Department at the University of Kansas, USA.