



Measuring Student Character in a Fair and Contextual Manner: Development of Local Culture-Based Instruments with the IRT Model

Vinsensius Bawa Toron^{1*}, Marzuki², Ali Muhson³, Florens Maxi Un Bria⁴, Hermania Bhoki⁵, Fidelis Regi Waton⁶

^{1,3}Program Studi Penelitian dan Evaluasi Pendidikan (PEP) Universitas Negeri Yogyakarta, Indonesia

¹Program Studi Pendidikan Keagamaan Katolik STP Reinha Larantuka, Indonesia

⁴Program Studi Pendidikan Keagamaan Katolik STIPAS Kupang, Indonesia

⁵Program Studi Pendidikan Keagamaan Katolik STP Reinha Larantuka, Indonesia

⁶Kölner Hochschule für Katholische Theologie, Germany

*Correspondence to: vinsensiusbawa.2022@student.uny.ac.id; vinsensius@stprenya-lrt.sch.id

Abstract: The importance of character education in the formation of a generation that is intellectually endowed but also morally and socially accountable should not be underrated. Nonetheless, evaluating student character remains a problem, especially across diverse local cultural backgrounds. In this research, an attempt was made to come up with a measure of character among junior high school students based on the local cultural values, and to achieve this purpose, the current psychometric theory of the item Response Theory (IRT) was used in the development of the instrument, as well as expert and data validation through interviews, questionnaires, and observations. The Generalised Partial Credit Model (GPCM) was the most suitable as determined in the analysis, and all the 68 items were highly conforming. The unidimensionality assumption was met (MSA = 0.977; first-factor variance = 77.57%), and the parameter correlations implied invariance and fairness across student groups. The instrument has a total information value of 403.08, a standard error of 0.05, and an information area probability of 98.52%, which indicates high precision and coverage of the measurement. Qualitative findings proved cultural relevance and semantic clarity. The study will contribute to the field of educational evaluation by offering a valid, reliable, and culture-adaptive tool that can serve as a guide for both research and research-practice in character education across diverse settings.

Keywords: assessment; item response model (irm); psychometrics; student personality; school culture

Recommended citation: Toron, V. B., Marzuki., Muhson, A., Bria, F. M. U., Bhoki, H., & Waton, F. R. (2026). Measuring Student Character in a Fair and Contextual Manner: Development of Local Culture-Based Instruments with the IRT Model. *Journal of Innovation in Educational and Cultural Research*, 7(2), 364-375.

INTRODUCTION

The concept of character education has gained a vital role for broad-minded persons in the 21st century (Ibrahim et al., 2024). It is becoming an increasingly important tool for responding to social and technological changes, yet its implementation in the curriculum poses a serious issue (Omodan et al., 2023). In the contemporary context of moral crises, social disintegration, individualism, and violence in educational settings, character education is seen as a key ingredient in developing a generation not only cognitively smart but also emotionally and spiritually tough. This highlights the fact that the education system should not merely be academic-based (Ribeiro et al., 2021) but also foster students' holistic nature and personality (Biswas & Verma, 2022). The government in Indonesia has responded to the character education emergency by implementing a number of policies, including the Movement to Strengthen Character Education and the introduction of Pancasila Student Profile values into the Independent Curriculum (Rahmawati et al., 2022). As the Ministry of Education and Culture (2021) states, character education is where the scheme for creating high-quality, internationally competitive Indonesian human resources is born (Asmara et al., 2025).

Practically, though, attempts to introduce character education generally face the problem of integrating it into the learning process (Veckalne & Tambovceva, 2022), of valid measuring student character (Syarnubi et al., 2021), and of pedagogical knowledge of the role of teachers as role models (Ekmekci & Serrano, 2022; König et al., 2023; Wraae et al., 2022). Suprpto et al. (2021) also note that one obstacle to empowering character education is the poor alignment between local values and the formal curriculum. Moreover, reflective and contextual pedagogies that align with local cultural values have not been fully understood (Dawson et al., 2022). Regarding the measurement of character, it is still unusual to develop a modern, psychometrically based measure that accounts for cross-cultural validity and local contexts (Ambuehl & Inauen, 2022; Lacko et al., 2022). In this regard, the role of junior high school teachers is strategic not only in providing the subject matter, but also as major players in the development of the character of students with different socioeconomic and cultural backgrounds (Sutama et al., 2021; Culp & Salvador, 2021; Martin et al., 2021; Yang et al., 2021). However, no deep empirical research has been conducted to assess the extent to which teachers can be

instrumental in fostering character among students, especially through educational psychology strategies and scientifically responsible evaluation frameworks. As a psychological phenomenon, character cannot be easily quantified by viewing it through surface-level eyes (Powell et al., 2021; Souissi et al., 2022; Uher, 2021). It has to be a scientific method that emphasizes validity and reliability, and it should incorporate psychometric techniques such as Item Response Theory (IRT) (Stănculescu, 2023). This method offers a more precise representation of students' inherent abilities or inclinations by examining how they respond to a developed instrument (Stephenson et al., 2020). IRT has benefits in assessing latent traits that cannot be directly observed, such as empathy, discipline, or tolerance (Henninger & Meiser, 2020; Nagy & Ulitzsch, 2022). When applied to character education, it helps differentiate between responses stemming from a person's character and those stemming from the situation.

It is strategically important because integrating an IRT-based mixed-methods approach will bring together two major strengths: the ability to understand the nature of social, cultural, and student life dynamics through qualitative approaches, and the possibility to assess character traits using an objective, standardized method. Mixed methods have been useful for creating and modifying measurement instruments, especially in behavioral research. To make the Illness Perception Questionnaire culturally sensitive and quantitatively accurate, Shiyabola et al. (2021) employed an exploratory sequential mixed-methods design; they used the qualitative results from focus group discussions to develop quantitative survey items. IRT is also instrumental in psychometric assessment, where it can provide a statistical model of how an individual responds to a psychological test item. As Lang and Tay (2021) note, the theory is used to explain the effects of different variables on the probability of a specific response and thus to provide a better representation of the construct being measured.

Recent studies show that IRT is optimizable to complex modeling situations. Henninger and Meiser (2020) place significant emphasis on the use of the divide-by-total model to understand response styles and the combination of several dimensions in response patterns. Such a methodology allows the researcher to obtain more insights into respondent variation and to increase the accuracy of psychological measurement. The qualitative and quantitative methods, combined with sophisticated IRT modeling, can create a psychometric instrument that is more culturally sensitive, conceptually correct, and reliable to apply across various populations. This would enhance the instrument's internal validity and make it more relevant to certain social and cultural settings, thereby improving its contribution to fairer and more meaningful measurement.

The study is a novelty, namely in creating and testing character education with junior high school students. The research does not just test character qualitatively but also structures and suggests a character testing tool using a contemporary psychometric methodology through IRT. Its value lies in its relative approach to three IRT models (GRM, GPCM, and PCM), with statistical measures (AID, BIC, and log-likelihood) used, a method that is uncommon in research examining the contextual and locally based values-based character education. At the conceptual and design levels, these tools are supposed to comply with the principles of unidimensionality and parameter invariance, which will help provide the same measure of student character across social situations. Similarly, when the test data functionality is intentionally examined, the tool is expected to exhibit high accuracy and low measurement error within the specified ability scope. Moreover, this paper reaffirms the relevance of teachers as psycho-pedagogical agents rather than curriculum implementers, who are expected to play an active role in shaping students (through instructional practices, role modeling, and emotional relationships), especially in unique sociocultural settings. The focus is to design a student character measurement tool based on educational psychology principles and validated using IRT to generate a valid and reliable measurement tool.

METHODS

The proposed research uses a mixed-methods design, combining quantitative and qualitative approaches (Adu et al., 2022). According to Foronda et al. (2022), it is stressed that, to offer a comprehensive picture of character education, evaluation tools should be designed to include aspects of local cultural values. The fact that the instruments have been developed taking into account the local cultural values not only strengthen the correspondence to the context, but also provides an opportunity to make sure that the process of evaluating character education reflects the identity, wisdom, and social norms existing in the communities where the education occurs Sharon E. et al. (2022); Tzima et al. (2020); Wright et al. (2021). The primary factor in the data collection and analysis process is teachers' active participation as research partners. The teacher is not only an object of study, but also the teacher is involved in the provision of input, testing, and reflection of the developed instruments, thus ensuring that the outputs can be more applicable and in correspondence to the actual learning context in the school (Jurkowski et al., 2023; Paju et al., 2022).

In the development and validation phases of the measurement instruments for character, purposive sampling methods were used to select a total of 529 students from junior high schools. The choice is made based on the representation of different school backgrounds in the East Nusa Tenggara Province of Indonesia (urban, rural, and remote areas) to ensure a variety of sociocultural settings. The educators were experimenting

with the instrument items, along with empirical validation, based on the principles of the Item Response Theory (IRT) approach, which corresponds to the provisions for creating the psychological instrument (Cruchinho et al., 2024) within the framework of validity and reliability of psychometric measuring tools.

The tool employed in the research is a questionnaire containing 50 statements on a 4-point Likert scale; this instrument will encompass elements related to character dimensions, including empathy, responsibility, discipline, and cooperation. This instrument was prepared in a participatory way through a focus group discussion (FGD) with teachers as partners, and the validation tests were conducted by experts in the education and psychology disciplines. This method adheres to the standard practice in psychometric instrument development, with emphasis on content validity and construction. The application of a 4-point Likert scale is also expected to reduce bias toward a neutral response and prompt the subject to adopt a more aggressive response (Xie & DeVellis, 1992). Also, the item selection and analysis are performed according to item response theory (IRT), which is one of the most popular methods in educational psychology to ensure item reliability and discrimination (Fayers, 2004). Lickona says that the construction principles of indicators and statement items are also made with respect to local culture and character values (Dodd, 1992). By doing this, the instrument is psychometrically sound, not just contextually and culturally relevant.

In this research, data collection was conducted triangulatively to achieve data validity and richness across different perspectives. Three major techniques are employed. The survey was carried out by distributing questionnaires to teachers at junior high schools, and the teachers were the main respondents. This questionnaire will include statements that evaluate students' character based on the teacher's direct observation during the learning and socializing processes at the school. As described by Creswell and Clark, this method allows researchers to obtain large volumes of quantitative data, which can be analysed using statistical methods, which fits the survey method of character education research (Bazeley, 2010; McMillan & Schumacher, 2010). Several teachers were interviewed openly and asked to complete questionnaires. The objective of such interviews is to obtain a more in-depth insight into how teachers perceive and notice students' character and attitudes, and into their perceptions of the clarity and validity of the items in the instruments. This method facilitates the qualitative aspect of a mixed research design, enabling researchers to study the social and cultural undertones that cannot be identified using questionnaires (Creswell, 2004). There were selective observations limited to learning situations and school activities, with the sole aim of pursuing further analysis of some instrument items found to be unsuitable in the IRT analysis. These observations aim to explain behavior that is not aligned with the instrument's theoretical requirements. The combination of observations, interviews, and questionnaires will make the results more reliable and enable the researchers to explain the differences in data more effectively (Denzin, 2012; Patton, 2002).

In this research, the quantitative analysis was conducted using Item Response Theory (IRT), which enables the quality of the instrument's items to be evaluated in detail. RStudio is the tool used to analyse the question items, and the analysis results are based on IRT. The most appropriate IRT models, including the Grade Response Model (GRM), Generalized Partial Credit Model (GPCM), and Partial Credit Model (PCM), are selected based on statistical indicators such as Akaike Information Criteria (AIC), Bayesian Information Criteria (BIC), and log-likelihood. To ensure the model fits the data, the choice of the latter is important, as highlighted in current psychometric research (Bürkner, 2021; Caycho-Rodríguez et al., 2022; Shim et al., 2022; Zielińska et al., 2022). Before implementing IRT, a unidimensional test was performed using Exploratory Factor Analysis (EFA), which targets the Measure of Sampling Adequacy dimension, in line with the unidimensional principle in IRT theory (de Barros Ahrens et al., 2020; Wang et al., 2021). Then, item-fit analysis was performed to test whether items fit or did not fit (misfit) the models. Parameters of discrimination and difficulty were studied to determine the degree to which they could distinguish people according to the level of character measured as written in the standard IRT practices to promote the fairness of the instrument, in line with the significance of the measurement invariance principle in the IRT literature (Mardapi & Kartowagiran, 2011; Pratama, 2020).

The transcript of the open-ended interview with the teacher was analyzed qualitatively, by using the thematic approach, to investigate how they perceive the character of the students, the connection between the measured values and the local cultural context, and the meaning of inappropriate items (Not Fit) in the outcome of the quantitative analysis. This method will be used to incorporate participants' lenses into the instrument revision process, so that the instrument development outcome is more contextual, meaningful, and sensitive to the local cultural values of the school community. The given thematic analysis is based on the method of assessing the conceptual validity of an instrument, in which the meanings participants apply to the narratives are understood (Nowell et al., 2017).

To provide the instrument with character evaluation validity and relevance, the use of local cultural values, and quantitative (questionnaires, IRT analysis) and qualitative (interviews, observations) data is interpreted and synthesized to ensure the relevance and validity of the tools. This is based on a mixed-methods approach to construct validity (Messick, 1995), as it evaluates the appropriateness of the behavioral indicator, avoids cultural bias, and addresses the affective, conative, and cultural aspects of students.

RESULT AND DISCUSSION

Table 1. Choose The Fit Model

Model	AIC	SABIC	HQ	BIC	Loglik	N Fit	N Not Fit
GRM	48577.26	48893.11	49058.76	49807.31	-24000.63	64	8
GPCM	49040.69	49356.55	49522.19	50270.74	-24232.35	65	7
PCM	49299.42	49537.40	49662.21	50226.22	-24432.71	60	12

Table 1 provides an assessment of three item response theory models applied in this study, namely the Graded Response Model (GRM), the Generalized Partial Credit Model (GPCM), and the Partial Credit Model (PCM). The models were evaluated to identify the best framework for measuring student character, based on teacher evaluations at junior high schools. The results show that the most suitable model in this case is the GRM, as the lowest AIC (Akaike Information Criterion) value of 48577.26 provides the optimal balance between model fit and complexity. This finding is also established by the minimum value of SABIC (Sample-size Adjusted BIC) of 48893.11 and HQ (Hannan-Quinn) of 49058.76. GRM also showed the lowest BIC (Bayesian Information Criterion) value (49807.31), indicating a less complex scenario than GPCM and PCM. Additionally, the large logLig (log-likelihood) value for the GRM indicates that this model provides the best fit to the character data, as evaluated by teachers. GRM was also found to be a stable instrument, and 64 items were adequate to fit the model, with only 8 items identified as outliers, which supports its performance in providing a more detailed description of teacher responses as a measure of student character.

In comparison with the prior research that typically assumed the use of IRT models like GPCM or PCM to procedural skills or generic instruments of attitude, this study shows that GRM is very much appropriate to the measurement of character, accommodating the complex and multidimensional nature of character in the education field, where earlier research has usually been more prone to using other IRT models. This research paper contributes to educational psychology and psychometric measurement by using empirical data on parenting to demonstrate that GRM can yield valid, reliable, and culturally relevant measures of student character. The result contributes to theoretical explanations of IRT by revealing that GRM applies to multidimensional character constructs. In practical terms, the findings give educationalists and scholars an approach to develop tools that are more appropriate for capturing teachers' judgments of students' attitudes and values and for enhancing the application of psychological testing strategies in the character education process.

It was found that the GPCM was the best-fitting model, with all items meeting the assumptions. The observation is consistent with previous studies (Bürkner, 2021; Dai et al., 2021) that highlight the strengths of the GPCM in modeling differences in item difficulty and sensitivity on Likert-type scales. Theoretically and empirically, the GPCM provides the flexibility to capture varying degrees of value internalization among students, particularly when the constructs are multiple, such as empathy, collaboration, and responsibility. These aspects are directly associated with regional cultural values: Gasik Arik (mutual respect), Tonga Danga (mutual honor), Pohen Haben (cooperation), and Tedun Toda (mutual assistance), which emphasize togetherness in life, generosity, and social sensitivity. In contrast, the GRM suggested that seven items (5, 10, 11, 18, 19, 35, 48) do not suit it. This can be explained by the fact that the model makes a strict assumption of linear and ordered gradation, which in many cases is disturbed by items (19, 35, 38, 40, 47, 48, 49), and the assumption that the partial credit across categories is equal does not reflect the variation of the expression of culture in the character of students. According to Islam and Ahmed (2021), PCM has proved more appropriate for the measurement of procedural skills, whose stages are well-defined and sequenced, as opposed to dynamic values based on social and cultural contexts. Such findings affirm that character is not a fixed psychological entity but is shaped by an individual's experience and the social-cultural context. Thus, models that are more flexible, like GPCM, are necessary for a valid assessment. The findings indicate that GPCM is the most representative model used to assess student character, with instruments designed to incorporate local cultural constructs.

The research is new because it uses the GPCM to measure character based on local cultural values. Other prior investigations have mostly used IRT models to assess generic attitudes or procedural skills. It is shown in this study how psychometric models can be softened and tailored to context-sensitive cultural environments, thus allowing a more sensitive and valid character assessment in emphasizing such cultural values as Gasik Arik (mutual respect), Tonga Danga (mutual honor), Pohen Haben (cooperation), and Tedun Toda (mutual assistance). Moreover, the study contributes to the evolution of the field of educational measurement, introducing a successful assessment model that allows incorporating local cultural values into contemporary psychometric practice. This kind of integration not only enriches the theoretical discourse of IRT-based approaches but also provides practical guidance on developing educational tools that are sensitive to students' social and cultural truths, as seen in Table 2.

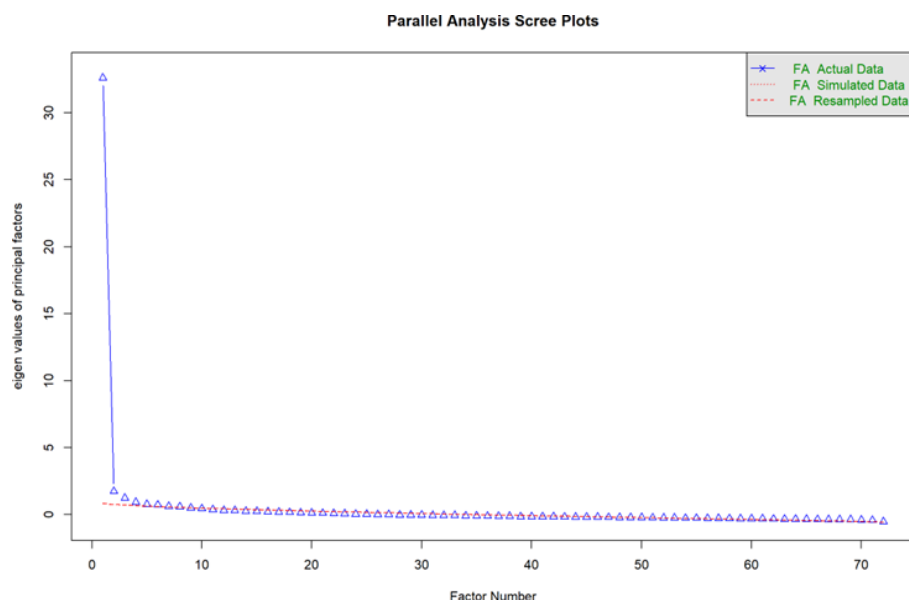
Table 2. The Fit Item in each Model

NO.	ITEM	GRM	GPCM	PCM	NO.	ITEM	GRM	GPCM	PCM
1	Fit	Fit	Fit	8	Notfit	Notfit	Notfit	Notfit	
2	Notfit	Notfit	Notfit	9	Fit	Fit	Fit	Fit	
3	Fit	Fit	Fit	10	Fit	Fit	Notfit	Notfit	
4	Fit	Fit	Notfit	11	Fit	Fit	Fit	Fit	
5	Fit	Fit	Notfit	12	Fit	Fit	Fit	Fit	
6	Fit	Fit	Notfit	13	Fit	Fit	Fit	Fit	
7	Fit	Fit	Fit	14					

Sources: Personal data (2025)

In this work, the assumption of unidimensionality, one of the conditions for the analysis under Item Response Theory (IRT), was evaluated using the Measure of Sampling Adequacy (MSA). The outcome revealed an MSA value of 0.977, which is significantly higher than the minimum value of 0.5. It means the sample size is adequate for an exploratory factor analysis (EFA) to determine the structural validity of the student character instrument items. A unidimensional ratio of 77.57 percent implies that there is one dominant factor (the first factor) that explains over 20 percent of the variation in the data. These results are in line with [McNeish and Wolf \(2023\)](#), who argued that when the ratio of the first factor's variance is greater than 20, the unidimensionality assumption is met. The scree plot further demonstrates a sharp decline in the first factor, cementing the conclusion that this is an instrument for measuring a single latent trait, student character. The same findings are also supported by the articles by [Imyati et al.](#) and [Lestari et al. \(2021\)](#), who developed a character instrument grounded in Pancasila values and used IRT analysis, which also showed a high unification ratio as a measure of construct validity.

Nevertheless, the findings differ slightly from those of [Parker et al. \(2023\)](#), who stated that some dimensions of students' character can be multidimensional, as different cultural settings and learning conditions can lead to variation. The IRT model employed thus needs to be adjusted (e.g., to a multidimensional IRT model). In this way, the unidimensionality assumption of the study was met, indicating that the constructed tool is valid enough to assess students' character in junior high schools using modern psychometric methodology (see [Figure 1](#)).

**Figure 1.** Screen Plot of EFA

Contrary to the previous findings in which partial multidimensionality was often reported in the assessment of character, this study provides significant evidence of unidimensionality in a character measure scale, which was created with references to local cultural values: Gasin Arik (mutual respect), Tonga Danga (mutual honor), Pohen Haben (cooperation), and Tedun Toda (mutual help). It is novel in that it shows that the measurement cannot be broken down into distinct dimensions due to the introduction of such contextualized cultural constructs, but can be combined into a single latent factor. This aligns with the psychometric thesis that culturally based character values can be effectively represented in a unidimensional IRT framework. The proposed research has theoretical value, as it reinforces the discussion concerning the applicability of IRT unidimensional models to highly complex, culturally entrenched constructs such as student character. It offers empirical support for the possibility of unidimensionality without compromising cultural particularity, thereby

closing a gap between psychometric rigor and education in the context. In practical terms, the results should provide educators and researchers with a sound methodological basis for designing effective, culturally relevant character assessment tools to facilitate more precise assessment of students' attitudes and values in both school and community settings.

Table 3. Invariant Interpretation

Correlation	P_Value	Interpretation
0.452	0	The discriminant parameter invariance assumption is met
0.315	0.007	The difficulty in the invariance assumption is met
0.961	0	The Ability invariance assumption is met

The results of the analysis presented in Table 3 indicate that the assumption of parameter invariance for discrimination, difficulty, and ability has been met. For the discrimination parameters, difficulty parameters, and ability parameters, the correlation values were 0.452 ($p = 0.000$), 0.315 ($p = 0.007$), and 0.961 ($p = 0.000$), respectively. This large p-value indicates that these parameters are consistent and do not differ across groups, suggesting that the character evaluation tool is impartial and can be used across different groups of students. These results align with other studies that examined the validity and reliability of the test instruments using Item Response Theory (IRT) by [Sinval et al. \(2022\)](#) and [Bader et al. \(2022\)](#), which found that the satisfaction of the invariance assumption is a measure of validity and reliability. The analysis also revealed that the mean ability (observation) of students was -0.337, and the mean measurement error was 0.372. The fact that the average scores are almost equal to zero means that most students are at the moderate, not very high or very low, level of ability, indicating an equal distribution of abilities in the population under investigation. Nevertheless, these findings may be considered a critique, as [Holzinger et al. \(2020\)](#) reported inconsistent findings. The average ability of students in character measurement may be prejudiced if teachers lack sufficient pedagogical and evaluation skills to comprehend the psychological aspects of character education.

What is unique about the current research is that it demonstrates that measurement invariance across IRT parameters can be achieved with tools that take into account local cultural values, including Gasing Arik (mutual respect), Tonga Danga (mutual honor), Pohen Haben (cooperation), and Tedun Toda (mutual help). In contrast to earlier studies, which mainly examined invariance in standardised or generic cognitive tests, this study also shows that fairness and stability across groups can be achieved in a complex arena like character education, where cultural interplay might be thought to introduce variability. This offers fresh perspectives on how flexible IRT-based invariance testing is for culturally encoded constructs. The study has contributed to research on the measurement and psychometrics of education by empirically demonstrating that an instrument constructed on cultural values can be assessed for rigorous invariance, thereby guaranteeing equity across a diverse student sample. These results, in theory, add to debates about the universality and flexibility of IRT models as applied to non-cognitive domains such as attitudes and character. In practice, the study offers learning institutions and policymakers strong methodological evidence that character-measurement instruments with cultural backgrounds can still be psychometrically fair and, hence, their application across various school environments. This lies between cultural relevance and the objectivity of measurement and helps further the discourse on both character education and contemporary psychometric practice.

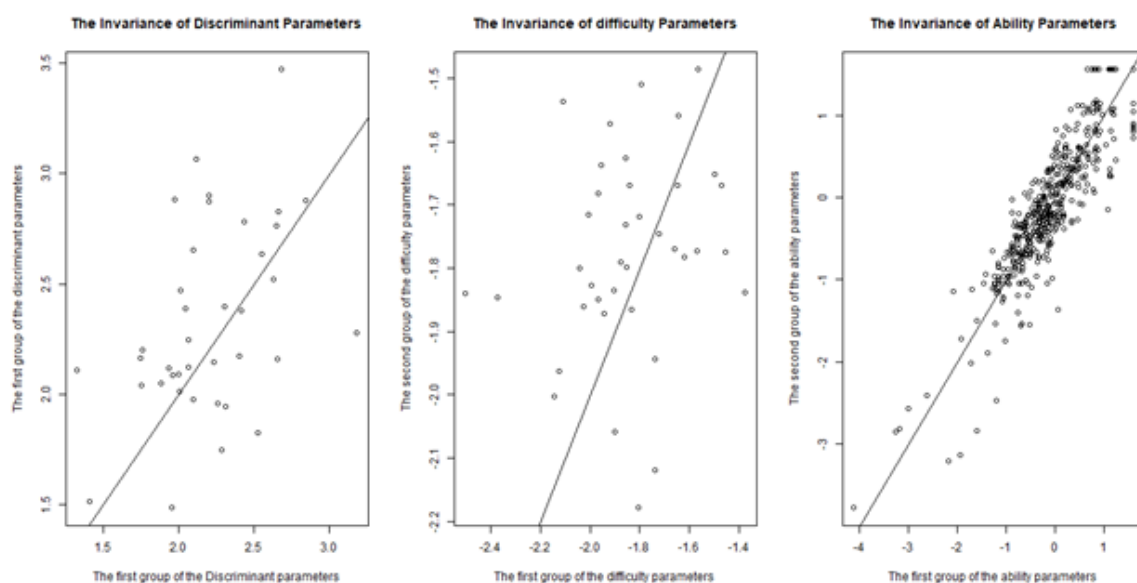


Figure 2. Parameter Invariant

Figure 2 shows that the analysis of the invariance parameter shows that the three primary parameters in the model, Discriminant, Difficulty, and Ability, satisfy the parameter determination assumption. This implies that the values of these parameters are stable and reliable across the diverse conditions tested. To be more precise, the Discriminant parameter is correlated with other variables or conditions ($r = 0.452$), but the p-value is 0.000, which is too low to conclude it is statistically insignificant. These results indicate that such parameters are not only constant but also sensitive to differentiate responses of different natures. There is also stability in the difficulty parameter, with a correlation of 0.315 and a p-value of 0.007, which again supports the assumption that the items' difficulty levels in the instrument are reasonable and relevant to the situation of the junior high school students under study. The ability parameter was very high, with a value of 0.961 and a p-value of 0.000, indicating that the student's ability, as a background factor and as a personality trait, did not change or fluctuate across the conditions in which they were tested.

The novelty of the current study is that it considers the analysis of invariance parameters in the measurement of students' character in relation to local cultural values, which has not been examined extensively in previous studies. In comparison to conventional models that simply test reliability and validity, the approach ensures the invariance of the parameters of discrimination, difficulty, and ability across various testing conditions, thereby rendering the instrument stronger and more contextually pertinent. The contribution and effects of these findings are immense. In practice, the findings offer educators and policymakers a valid and culturally grounded instrument to better and more objectively gauge the character of junior high school students. In theory, the study would improve the measurement model by reinforcing the empirical evidence that parameter invariance would improve the reliability and contextual sensitivity of the character education assessment.

Table 4. Information Function Interpretation

Interpretation	Numeric	Value
The Intersection of information function and error Standard on the left	-5.176	1.006
The Intersection of Information Function and Error Standard on the Right	2.405	1.01
The Total Information Value of All Items		403.08
The Error Standard of All Items		0.05
The Probability of the Shaded Area		98.52 %
The instrument is Suitable For measuring ability at intervals	-4	2.405

Table 4 demonstrates the interaction of the function with the standard of error on both ends of the capability spectrum. The degenerative point at the junction of the information work and the error is within the range of 5.179 to 1.006 on the left side, indicating that the instrument is well scaled for people with low- to medium-ability. The gap between 2.0405 and 1.01 on the right indicates that the tool can still provide credible information to competent individuals. Such a broad scope indicates that the tool is evenly balanced in the information it covers and can adequately address differences in students' abilities. The overall information value of 403.08 is the sum of all item information in the instrument and indicates that the instrument successfully yields accurate capability estimates. The greater the total value of the information, the more precise the instrument will be in identifying the respondent's level of ability. The instrument has been supported by a low standard measurement error of 0.05 and has been shown to possess high precision. The low standard error supports the reliability of any estimate of individual capacities derived using this instrument.

The innovative aspect of the current study is the analysis of invariance parameters and information functions in the measurement of student character using measurement tools in accordance with local cultural values. This is a two-sided approach that has hardly been applied in past studies of character education, which have only addressed classical reliability and validity. This research is invariant and prescriptive and does not require any adjustments, unlike traditional models, ensuring that the instrument will be fair across groups and very accurate in measuring a wide range of ability levels. The contribution and influence of such findings are enormous. In a practical sense, the findings can provide educators, curriculum developers, and policymakers with a powerful, culturally grounded instrument to assess the character of junior high school students more accurately and fairly. In terms of theory, the research study will help develop the Items Response Theory (IRT)-based assessment as it will provide empirical findings that invariance testing and item analysis through information function can improve the equity, accuracy and situational sensitivity of character assessment scales.

This research study has demonstrated that the character measurement tool developed possesses good psychometric characteristics, with an information coverage percentage of 98.52, as indicated by the information function graph. This implies that almost all important sections are well represented in the measurement, providing precision and fairness across different levels of student capability. This kind of evidence is important in proving that the instrument works best at certain ability levels, minimizing the error and bias in evaluating the character. In comparison to previous research, which in many cases concentrated on indices of general validity and reliability (Ahmed & Ishtiaq, 2021), the study offers methodological novelty by using the information function in Item Response Theory (IRT) to determine the best ability intervals. This has been supported by

recent studies, including Hur and Seo (2023), who used a polytomous IRT model to optimize character quality tests in medical education, and Toraman et al. (2022), who emphasized the importance of measurement invariance to provide equity in group-based distance education assessment. The multidimensional IRT approach to complex response structures, as shown by Nie et al. (2024), demonstrates the methodological progression of our approach to multidimensional character traits and thus is open to expansion.

The findings are also very much in line with the goals of this study, namely to develop a fair, contextual, and dependable character measurement tool. The fact that the measurement coverage is almost complete means that the low- and high-ability students are being assessed on a fair basis, which is consistent with the concept of cultural and educational psychology. This, in practical terms, has offered teachers and policymakers a strong instrument for building student character. In principle, it will extend the scope of IRT applications beyond the realm of cognitive assessment into the affective and value domains, thereby enriching current research in psychometrics and providing a firmer foundation for character education assessment.

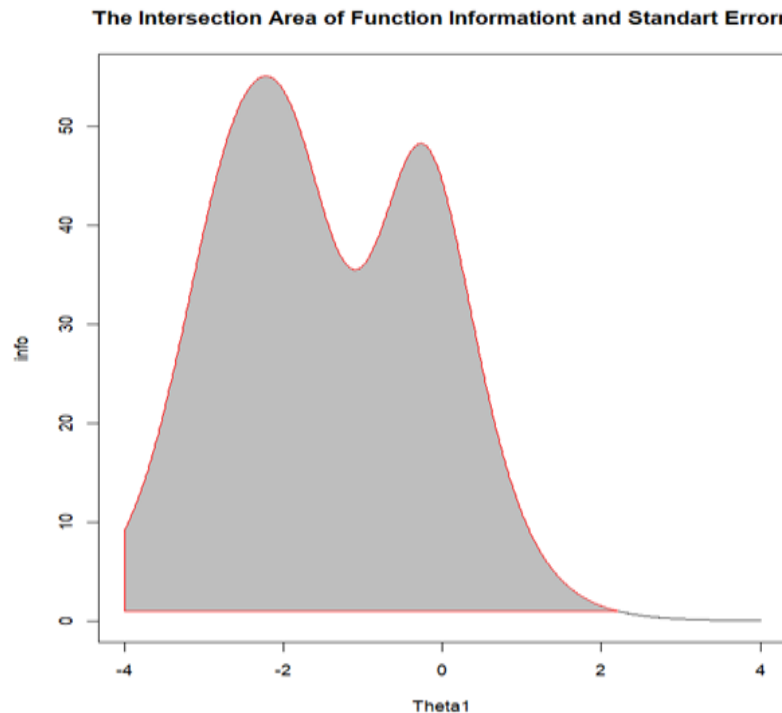


Figure 3. Information function

CONCLUSION

This research study has demonstrated that the character measurement tool developed possesses good psychometric characteristics, with an information coverage percentage of 98.52, as indicated by the information function graph. This implies that almost all important sections are well represented in the measurement, providing precision and fairness across different levels of student capability. This kind of evidence is important in proving that the instrument works best at certain ability levels, minimizing the error and bias in evaluating the character. The findings are also very much in line with the goals of this study, that is, to come up with a fair, contextual, and dependable character measurement tool. The fact that the measurement coverage is almost complete means that the low- and high-ability students are being assessed on a fair basis, which is consistent with the concept of cultural and educational psychology. This, in practical terms, has offered teachers and policymakers a strong instrument for building student character. In principle, it will extend the scope of IRT applications beyond the realm of cognitive assessment into the affective and value domains, thereby enriching current research in psychometrics and providing a firmer foundation for character education assessment.

ACKNOWLEDGEMENTS

The author admits that the artificial intelligence systems of Paperpal and Grammarly were used to assist in the writing of this article, especially in areas such as grammar checkers, sentence structure, text coherence, and academic style. The guidelines applied are grammar editing, sentence rearrangement to make it more comprehensible, and enhancing the academic viability of the manuscript. The result of these instructions is that they help improve and refine the quality of the writing without altering its content or the underlying meaning. Even though the author does not deny using this AI-based tool, they claim that all ideas, analyses, arguments, and interpretations in the current article are entirely their own.

REFERENCES

- Adu, J., Owusu, M. F., Martin-Yeboah, E., Gavidia, L. A. P., & Gyamfi, S. (2022). A Discussion of Some Controversies in Mixed Methods Research for Emerging Researchers. *Methodological Innovations*, 15(3), 321–330.
- Ahmed, I., & Ishtiaq, S. (2021). Reliability and Validity: Importance in Medical Research. *JPMA. The Journal of the Pakistan Medical Association*, 71(10), 2401–2406.
- Ambuehl, B., & Inauen, J. (2022). Contextualized Measurement Scale Adaptation: A 4-Step Tutorial for Health Psychology Research. *International Journal of Environmental Research and Public Health*, 19(19), 12775.
- Anif, S., Prayitno, H. J., Narimo, S., Fuadi, D., Sari, D. P., & Adnan, M. (2021). Metacognition of Junior High School Students in Mathematics Problem Solving Based on Cognitive Style. *Asian Journal of University Education*, 17(1), 134–144.
- Asmara, O. J., Morawska, A., Hoang, A., & Riany, Y. E. (2025). How Far Does the Apple Fall From the Tree? A Cross-Sectional Study of the Relationship Between Child and Parent Self-Regulation Mediated by Parenting Practices. *International Journal of Behavioral Development*, 01650254251327364.
- Bader, M., Horsten, L. K., Hilbig, B. E., Zettler, I., & Moshagen, M. (2022). Measuring the Dark Core of Personality in German: Psychometric Properties, Measurement Invariance, Predictive Validity, and Self-Other Agreement. *Journal of Personality Assessment*, 104(5), 660–673.
- Bazeley, P. (2010). Book Review: V. L. Plano Clark and J. W. Creswell (Eds.) *The Mixed Methods Reader*. *Journal of Mixed Methods Research*, 4(1), 79–81.
- Biswas, A., & Verma, R. K. (2022). Engine of Entrepreneurial Intentions: Revisiting Personality Traits With Entrepreneurial Education. *Benchmarking: An International Journal*, 29(6), 2019–2044.
- Bürkner, P. C. (2021). Bayesian Item Response Modeling in R With brms and Stan. *Journal of Statistical Software*, 100(5).
- Caycho-Rodríguez, T., Vilca, L. W., Carbajal-León, C., White, M., Vivanco-Vidal, A., Saroli-Araníbar, D., Peña-Calero, B. N., & Moreta-Herrera, R. (2022). Coronavirus Anxiety Scale: New Psychometric Evidence for the Spanish Version Based on CFA and IRT Models in a Peruvian Sample. *Death Studies*, 46(5), 1090–1099.
- Creswell, J. W. (2004). Designing a Mixed Methods Study in Primary Care. *The Annals of Family Medicine*, 2(1), 7–12.
- Cruchinho, P., López-Franco, M. D., Capelas, M. L., Almeida, S., Bennett, P. M., Miranda Da Silva, M., Teixeira, G., Nunes, E., Lucas, P., & Gaspar, F. (2024). Translation, Cross-Cultural Adaptation, and Validation of Measurement Instruments: A Practical Guideline for Novice Researchers. *Journal of Multidisciplinary Healthcare*, 17, 2701–2728.
- Culp, M. E., & Salvador, K. (2021). Music Teacher Education Program Practices: Preparing Teachers to Work With Diverse Learners. *Journal of Music Teacher Education*, 30(2), 51–64.
- Dai, S., Vo, T. T., Kehinde, O. J., He, H., Xue, Y., Demir, C., & Wang, X. (2021). Performance of Polytomous IRT Models With Rating Scale Data: An Investigation Over Sample Size, Instrument Length, and Missing Data. *Frontiers in Education*, 6, 721963.
- Dawson, J., Laccos-Barrett, K., Hammond, C., & Rumbold, A. (2022). Reflexive Practice as an Approach to Improve Healthcare Delivery for Indigenous Peoples: A Systematic Critical Synthesis and Exploration of the Cultural Safety Education Literature. *International Journal of Environmental Research and Public Health*, 19(11), 6691.
- de Barros Ahrens, R., da Silva Lirani, L., & de Francisco, A. C. (2020). Construct Validity and Reliability of the Work Environment Assessment Instrument WE-10. *International Journal of Environmental Research and Public Health*, 17(20), 7364.
- Denzin, N. K. (2012). Triangulation 2.0. *Journal of Mixed Methods Research*, 6(2), 80–88.
- Dimiyati, K., Nashir, H., Elviandri, E., Absori, A., Wardiono, K., & Budiono, A. (2021). Indonesia as a Legal Welfare State: A Prophetic-Transcendental Basis. *Heliyon*, 7(8), e07865.
- Dodd, A. W. (1992). Educating for Character: How Our Schools Can Teach Respect and Responsibility. By Thomas Lickona. *NASSP Bulletin*, 76(545), 119–120.

- Ekmekci, A., & Serrano, D. M. (2022). The Impact of Teacher Quality on Student Motivation, Achievement, and Persistence in Science and Mathematics. *Education Sciences*, 12(10), 649.
- Fayers, P. (2004). Item Response Theory for Psychologists. *Quality of Life Research*, 13(3), 715–716.
- Foronda, C., Prather, S., Baptiste, D. L., & Luctkar-Flude, M. (2022). Cultural Humility Toolkit. *Nurse Educator*, 47(5), 267–271.
- Henninger, M., & Meiser, T. (2020). Different Approaches to Modeling Response Styles in Divide-by-Total Item Response Theory Models (Part 1): A Model Integration. *Psychological Methods*, 25(5), 560–576.
- Holzinger, A., Lettner, S., Steiner-Hofbauer, V., & Melser, M. C. (2020). How to Assess? Perceptions and Preferences of Undergraduate Medical Students Concerning Traditional Assessment Methods. *BMC Medical Education*, 20(1), 312.
- Hur, Y., & Seo, D. G. (2023). Development of a Character Qualities Test for Medical Students in Korea Using Polytomous Item Response Theory and Factor Analysis: A Preliminary Scale Development Study. *Journal of Educational Evaluation for Health Professions*, 20, 20.
- Ibrahim, N. N., Naidu, N. B. M., Shegaram, P. C., & Ahmad, N. S. (2024). Character Education in the 21st Century: A Bibliometric Analysis on 2000-2024 Scopus Database. *International Journal of Modern Education*, 6(22), 741–757.
- Islam, M. N., & Ahmed, D. H. (2021). Delaying the Temperature Fluctuations Through PCM Integrated Building Walls—Room Conditions, PCM Placement, and Temperature of the Heat Sources. *Energy Storage*, 3(5), e245.
- Jurkowski, S., Ulrich, M., & Müller, B. (2023). Co-Teaching as a Resource for Inclusive Classes: Teachers' Perspectives on Conditions for Successful Collaboration. *International Journal of Inclusive Education*, 27(1), 54–71.
- König, C. M., Karrenbauer, C., & Breitner, M. H. (2023). Critical Success Factors and Challenges for Individual Digital Study Assistants in Higher Education: A Mixed Methods Analysis. *Education and Information Technologies*, 28(4), 4475–4503.
- Lacko, D., Čeněk, J., Točík, J., Avsec, A., Đorđević, V., Genc, A., Haka, F., Šakotić-Kurbalija, J., Mohorić, T., Neziri, I., & Subotić, S. (2022). The Necessity of Testing Measurement Invariance in Cross-Cultural Research: Potential Bias in Cross-Cultural Comparisons With Individualism–Collectivism Self-Report Scales. *Cross-Cultural Research*, 56(2–3), 228–267.
- Lang, J. W. B., & Tay, L. (2021). The Science and Practice of Item Response Theory in Organizations. *Annual Review of Organizational Psychology and Organizational Behavior*, 8(1), 311–338.
- Lestari, L. (2021). *Pembentukan Akhlakul Karimah dalam Buku Pendidikan Karakter Islam Karya Dr. Marzuki, M. Ag* (Doctoral dissertation, Universitas Islam Negeri Sultan Syarif Kasim Riau).
- Mardapi, D., & Kartowagiran, B. (2011). Pengembangan Instrumen Pengukur Hasil Belajar Nirbias dan Terskala Baku. *Jurnal Penelitian dan Evaluasi Pendidikan*, 15(2), 326–341.
- Martin, C. S., Ramirez, C., Calvo, R., Muñoz-Martínez, Y., & Sharma, U. (2021). Chilean Teachers' Attitudes Towards Inclusive Education, Intention, and Self-Efficacy to Implement Inclusive Practices. *Sustainability*, 13(4), 1–17.
- McMillan, J. H., & Schumacher, S. (2010). *Research in Education: Evidence-Based Inquiry*, MyEducationLab Series. Pearson.
- McNeish, D., & Wolf, M. G. (2023). Dynamic Fit Index Cutoffs for Confirmatory Factor Analysis Models. *Psychological Methods*, 28(1), 61–88.
- Nagy, G., & Ulitzsch, E. (2022). A Multilevel Mixture IRT Framework for Modeling Response Times as Predictors or Indicators of Response Engagement in IRT Models. *Educational and Psychological Measurement*, 82(5), 845–879.
- Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic Analysis: Striving to Meet the Trustworthiness Criteria. *International Journal of Qualitative Methods*, 16(1), 1609406917733847.
- Omodan, B. I., Mpiti, P. T., & Mtsi, N. (2023). A Conceptual Analysis of What It Means to Decolonize the Curriculum. *International Journal of Learning, Teaching and Educational Research*, 22(7), 344–355.
- Paju, B., Kajamaa, A., Pirttimaa, R., & Kontu, E. (2022). Collaboration for Inclusive Practices: Teaching Staff Perspectives From Finland. *Scandinavian Journal of Educational Research*, 66(3), 427–440.

- Parker, L. J., Marx, K. A., Nkimbeng, M., Johnson, E., Koeuth, S., Gaugler, J. E., & Gitlin, L. N. (2023). It's More Than Language: Cultural Adaptation of a Proven Dementia Care Intervention for Hispanic/Latino Caregivers. *The Gerontologist*, 63(3), 558–567.
- Patton, M. Q. (2002). Two Decades of Developments in Qualitative Inquiry: A Personal, Experiential Perspective. *Qualitative Social Work*, 1(3), 261–283.
- Powell, T., Wegmann, K. M., & Backode, E. (2021). Coping and Post-Traumatic Stress in Children and Adolescents After an Acute Onset Disaster: A Systematic Review. *International Journal of Environmental Research and Public Health*, 18(9), 4865.
- Pratama, D. (2020). Analisis Kualitas Tes Buatan Guru Melalui Pendekatan Item Response Theory (IRT) Model Rasch. *Tarbawy: Jurnal Pendidikan Islam*, 7(1), 61–70.
- Rahmawati, Y., Taylor, E., Taylor, P. C., Ridwan, A., & Mardiah, A. (2022). Students' Engagement in Education as Sustainability: Implementing an Ethical Dilemma-STEAM Teaching Model in Chemistry Learning. *Sustainability*, 14(6), 3554.
- Ribeiro, D. L., Costa, M., Helmich, E., Jaarsma, D., & De Carvalho-Filho, M. A. (2021). 'I Found Myself a Despicable Being!': Medical Students Face Disturbing Moral Dilemmas. *Medical Education*, 55(7), 857–871.
- Sharon E., C., Lauren J., J., Jordan R., C., Abby A., K., Sharon K., P., Melody, R., Michele, K.-F., Negar, G., & Regine Beliard, V. (2022). A Systematic Review of Global Health Assessment for Education in Healthcare Professions. *Annals of Global Health*, 88(1), 1.
- Shim, H., Bonifay, W., & Wiedermann, W. (2022). Parsimonious Asymmetric Item Response Theory Modeling With the Complementary Log-Log Link. *Behavior Research Methods*, 55(1), 200–219.
- Shiyanbola, O. O., Rao, D., Bolt, D., Brown, C., Zhang, M., & Ward, E. (2021). Using an Exploratory Sequential Mixed Methods Design to Adapt an Illness Perception Questionnaire for African Americans With Diabetes: The Mixed Data Integration Process. *Health Psychology and Behavioral Medicine*, 9(1), 796–817.
- Sinval, J., Vazquez, A. C. S., Hutz, C. S., Schaufeli, W. B., & Silva, S. (2022). Burnout Assessment Tool (BAT): Validity Evidence From Brazil and Portugal. *International Journal of Environmental Research and Public Health*, 19(3), 1344.
- Souissi, S., Chamari, K., & Bellaj, T. (2022). Assessment of Executive Functions in School-Aged Children: A Narrative Review. *Frontiers in Psychology*, 13, 991699.
- Stănculescu, E. (2023). The Bergen Social Media Addiction Scale Validity in a Romanian Sample Using Item Response Theory and Network Analysis. *International Journal of Mental Health and Addiction*, 21(4), 2475–2492.
- Stephenson, C. R., Bonnes, S. L., Sawatsky, A. P., Richards, L. W., Schleck, C. D., Mandrekar, J. N., Beckman, T. J., & Wittich, C. M. (2020). The Relationship Between Learner Engagement and Teaching Effectiveness: A Novel Assessment of Student Engagement in Continuing Medical Education. *BMC Medical Education*, 20(1), 403.
- Suprpto, N., Prahani, B. K., & Cheng, T. H. (2021). Indonesian Curriculum Reform in Policy and Local Wisdom: Perspectives From Science Education. *Jurnal Pendidikan IPA Indonesia*, 10(1), 69–80.
- Syarnubi, S., Mansir, F., Purnomo, M. E., Harto, K., & Hawi, A. (2021). Implementing Character Education in Madrasah. *Jurnal Pendidikan Islam*, 7(1), 77–94.
- Toraman, Ç., Karadağ, E., & Polat, M. (2022). Validity and Reliability Evidence for the Scale of Distance Education Satisfaction of Medical Students Based on Item Response Theory (IRT). *BMC Medical Education*, 22(1), 94.
- Tzima, S., Styliaras, G., & Bassounas, A. (2020). Revealing Hidden Local Cultural Heritage Through a Serious Escape Game in Outdoor Settings. *Information*, 12(1), 10.
- Uher, J. (2021). Psychometrics Is Not Measurement: Unraveling a Fundamental Misconception in Quantitative Psychology and the Complex Network of Its Underlying Fallacies. *Journal of Theoretical and Philosophical Psychology*, 41(1), 58–84.
- Veckalne, R., & Tambovceva, T. (2022). The Role of Digital Transformation in Education in Promoting Sustainable Development. *Virtual Economics*, 5(4), 65–86.

- Wang, D., Zhu, F., Xi, S., Niu, L., Tebes, J. K., Xiao, S., & Yu, Y. (2021). Psychometric Properties of the Multidimensional Scale of Perceived Social Support (MSPSS) Among Family Caregivers of People With Schizophrenia in China. *Psychology Research and Behavior Management*, 14, 1201–1209.
- Wraae, B., Brush, C., & Nikou, S. (2022). The Entrepreneurship Educator: Understanding Role Identity. *Entrepreneurship Education and Pedagogy*, 5(1), 3–35.
- Wright, M., Getta, A., Green, A., Kickett, U., Kickett, A., McNamara, A., McNamara, U., Newman, A., Pell, A., Penny, A., Wilkes, U., Wilkes, A., Culbong, T., Taylor, K., Brown, A., Dudgeon, P., Pearson, G., Allsop, S., Lin, A., ... O'Connell, M. (2021). Co-Designing Health Service Evaluation Tools That Foreground First Nation Worldviews for Better Mental Health and Wellbeing Outcomes. *International Journal of Environmental Research and Public Health*, 18(16), 8555.
- Xie, Y., & DeVellis, R. F. (1992). Scale Development: Theory and Applications. *Contemporary Sociology*, 21(6), 876.
- Yang, Y., Li, G., Su, Z., & Yuan, Y. (2021). Teacher's Emotional Support and Math Performance: The Chain Mediating Effect of Academic Self-Efficacy and Math Behavioral Engagement. *Frontiers in Psychology*, 12, 651608.
- Zielińska, A., Lebuda, I., & Karwowski, M. (2022). Scaling the Creative Self: An Item Response Theory Analysis of the Short Scale of Creative Self. *Creativity Research Journal*, 34(4), 431–444.