Enhancing Non-Cognitive Assessments with GPT: Innovations in Item Generation and Translation for the University Belonging Questionnaire

Mingfeng Xue MINGFENGXUE@BERKELEY.EDU
Yunting Liu YUNTING99@BERKELEY.EDU
University of California, Berkeley
HuaXia Xiong HUAXIA$_b$*nu*@163.*com*
BeijingNormalUniversity

Editor: DONGLIANG GUO

Abstract

This study explores the application of GPT-3.5-turbo for item generation and translation in non-cognitive educational assessments, specifically focusing on the University Belonging Questionnaire (UBQ). The UBQ, designed to measure university students' sense of belonging across three dimensions, was expanded to include a new dimension on peer relationships, and translated into Chinese using GPT-3.5-turbo. A total of 25 new items, including those for the new dimension were generated and translated into Chinese, out of which 14 items passed the expert review. Psychometric analyses of the expanded and translated UBQ were conducted to evaluate reliability, internal structure, and external validity. The results demonstrate that the UBQ, with its new and translated items, maintains strong reliability and satisfactory internal structure, although the new items introduced some noise. Correlation analyses with the general belongingness scale revealed moderate associations with the acceptance dimension but weak associations with the overall scale. The study highlights GPT's potential in efficiently expanding and translating non-cognitive assessment tools. This work addresses crucial needs in educational assessments and provides a foundation for future advancements in item development and translation.

Keywords: Item Generation, Item Translation, University Belonging Questionnaire, Non-Cognitive Assessment, Large Language Model

1. Introduction

The growing integration of large foundation models, such as GPT-3.5, into educational assessment offers significant potential to streamline traditionally labor-intensive tasks like item generation and translation (Brown, 2020; Devlin, 2018; Wilson, 2023). While these models have been extensively studied for cognitive assessments, their application to non-cognitive measures remains relatively unexplored (Gierl and Haladyna, 2013; Kurdi et al., 2020). This gap is particularly important as non-cognitive factors, such as students' sense of belonging, are critical for predicting academic success and well-being (Slaten et al., 2018), and also non-cognitive factors are by nature less well defined as compared to cognitive factors thus posing more challenges on creating a sound measurement tool(Humphries and Kosse, 2017).

This paper investigates the use of GPT-3.5-turbo in automating the generation and translation of items for the University Belonging Questionnaire (UBQ), a tool designed to

measure students' sense of belonging within an academic institution (Slaten et al., 2018). We evaluate the model's effectiveness in generating new items, translating English items into Chinese, and focus on the evaluation from the psychometric perspective. Using the processes illustrated in Figure 1, the key research questions for this work are:

- **RQ1:** Do non-cognitive assessment items generated by GPT preserve satisfactory psychometric properties(i.e., reliability and validity) as established in the original UBQ tool?
- **RQ2:** Do non-cognitive assessment items translated by GPT have cultural adaptability, and if so, to what extent?

2. Related Work

2.1. University Belonging Questionnaire

Belongingness, recognized as a fundamental human need, is regarded as a deeply ingrained motivation for individuals (Allen et al., 2022). In educational settings, a sense of belonging has been shown to play a crucial role as a non-cognitive construct that positively contributes to students' success (Vargas-Madriz and Konishi, 2021). The UBQ was developed to measure university students' sense of belonging, providing a standardized tool for use across studies and disciplines in higher education (Slaten et al., 2018). The UBQ employs a four-point Likert scale ranging from *strongly disagree* to *strongly agree* and consists of 24 items, as detailed in **Appendix A**. Analyzed by explanatory factor analysis, the questionnaire can be divided into three dimensions: university affiliation (12 items), university support and acceptance (8 items), and faculty and staff relations (4 items). The total UBQ had an internal consistency of $\alpha = 0.94$, the three subscales had $\alpha = 0.92$ for university affiliation, 0.85 for university support and acceptance, and 0.88 for faculty and staff relations. Three subscales showed small to moderate correlations: university affiliation with university support and acceptance (r=0.65), university affiliation with faculty and staff relations (r=0.47), and university support and acceptance with faculty and staff relations (r=0.48) (Slaten et al., 2018).

The UBQ is selected for this study for several reasons: (1) the sense of belonging is an essential non-cognitive factor in educational environments, making the UBQ a widely used tool (Ahn and Davis, 2023); (2) there is a strong need for further item development, particularly regarding the faculty and staff relations dimension; (3) the absence of a dimension addressing peer relationships, which are critical in shaping students' sense of belonging (Gowing, 2019); and (4) the UBQ, originally developed in English, requires translation into other languages (e.g., Chinese) to extend its applicability to broader populations.

2.2. Automatic item generation

Items form the backbone of educational assessments, playing a pivotal role in evaluating knowledge and skills. Practitioners increasingly face pressure to expand item banks to ensure test security, accommodate diverse respondent populations, and keep pace with social and technological changes (Kaat et al., 2019; Xie et al., 2014). However, item development remains a costly, time-intensive, and labor-heavy process (Wilson, 2023; Lane et al.,

2016). Consequently, automatic item generation (AIG) has emerged as a promising solution for scaling item production (Gierl and Haladyna, 2013). Common AIG methods include syntactic and semantic approaches, rule-based systems, templates, and statistical methods (Gierl and Haladyna, 2013; Yao et al., 2012). Template-based AIG, for example, may keep the core structure of an item while altering key terms related to the measured construct.

As a natural language processing (NLP) task, AIG has seen the increasing application of large language models (LLMs) in educational assessments, yielding promising results in generating items, particularly for cognitive assessments (Laverghetta Jr and Licato, 2023; Attali et al., 2022; Lee et al., 2023; Bhandari et al., 2024). The majority of this research focuses on generating cognitive items, while the generation of items for non-cognitive constructs, such as social-emotional skills and attitudes (e.g., sense of belonging), remains underexplored (Farruggia et al., 2018). Yet constructing a non-cognitive assessment might be a harder problem since there is usually no clear structure and standardization for the factors being measured(Humphries and Kosse, 2017). Given the well-documented impact of non-cognitive factors on students' future success (Kautz et al., 2014; Carneiro et al., 2007), there is a pressing need for more research into non-cognitive item generation using LLMs. Furthermore, examining the psychometric properties of generated items is critical, as valid measurement depends on the coordination of multiple items or even multiple dimensions.

2.3. Item translation

Item translation is essential for assessments administered to respondents who speak languages other than the one used during item development. For example, most items for the Programme for International Student Assessment (PISA) are initially developed in English, requiring extensive translation efforts to support the international scope of the assessment (OECD, 2024). While LLMs have demonstrated strong translation capabilities (Radford et al., 2019), translation for educational assessments demands a higher standard of accuracy and cultural adaptation.

Traditionally, forward and backward translation has been the gold standard for cross-cultural item translation (Brislin, 1970). This method involves translating items from the source language into the target language, followed by a back-translation into the original language by an independent translator. Comparing the back-translation with the original text allows for the identification of discrepancies, which can then be resolved through revisions. Despite its widespread use, this method is time-consuming and costly, as it often requires professional translators. Moreover, the quality of translations depends heavily on the translator's skill, motivation, and attention to detail (Ozolins et al., 2020). In this context, LLMs offer the potential for efficient, contextually accurate translations, providing a scalable alternative to traditional methods.

3. Methods

3.1. Item Generation and Translation

We employed GPT-3.5-turbo for both item generation and translation tasks, as it was the most up-to-date version of GPT at the time of the study. For item generation, we used the default settings for temperature and Top P. The system component setup, shown in

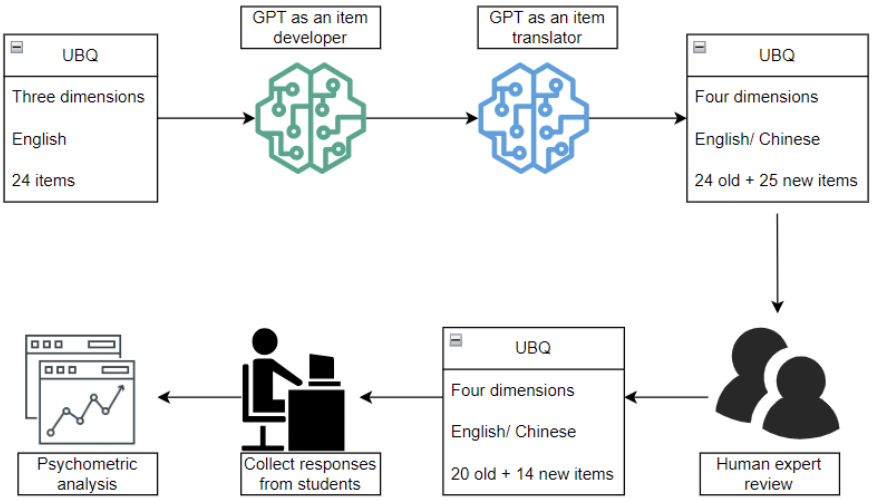

Figure 1: The process adopted in this research to expand and translate the UBQ

Figure 2, included specifying the role of GPT, defining of meaning of the four dimensions, and providing guidelines for item generation. Using the prompt template shown in Figure 3, GPT was instructed to generate five items for each existing dimension and 10 items for the new peer relationship dimension. The prompt included the 24 existing items to support few-shot learning and reduce the likelihood of duplicate items. In total, GPT generated 25 new items on four dimensions.

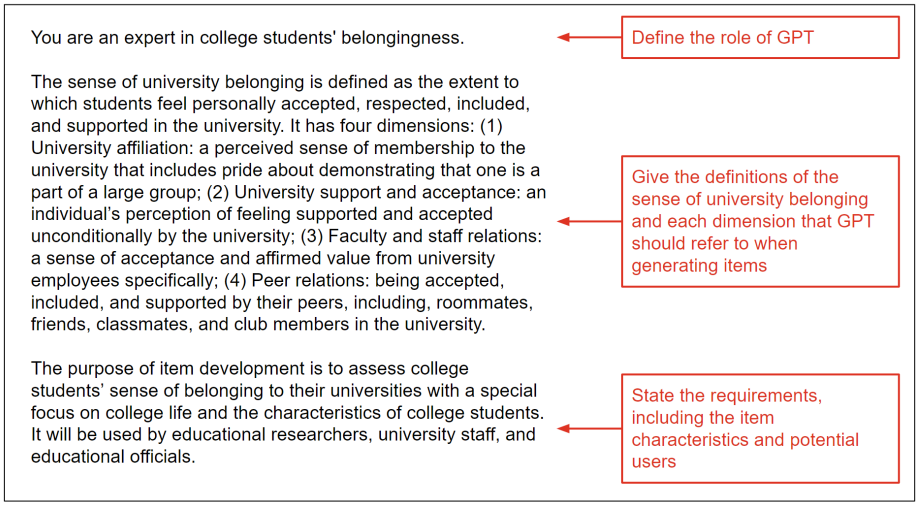

Figure 2: System instruction for GPT as an item generator

After generating the items and conducting an initial expert review, we used GPT to translate all items into Chinese. For this task, we defined GPT's role based on the criteria we use when hiring human translators in prior projects (see Figure 4). To leverage few-

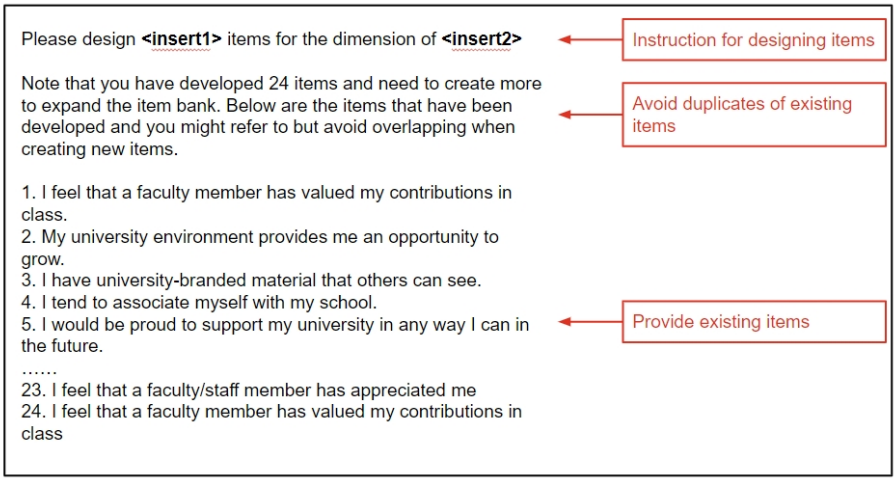

Figure 3: An item generation prompt template

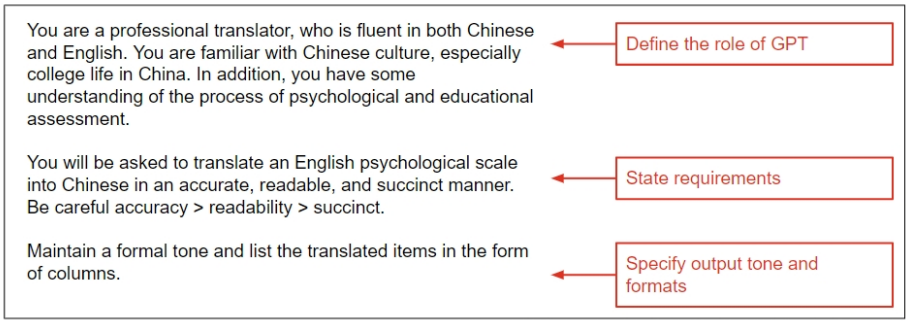

Figure 4: System instruction for GPT as an item translator

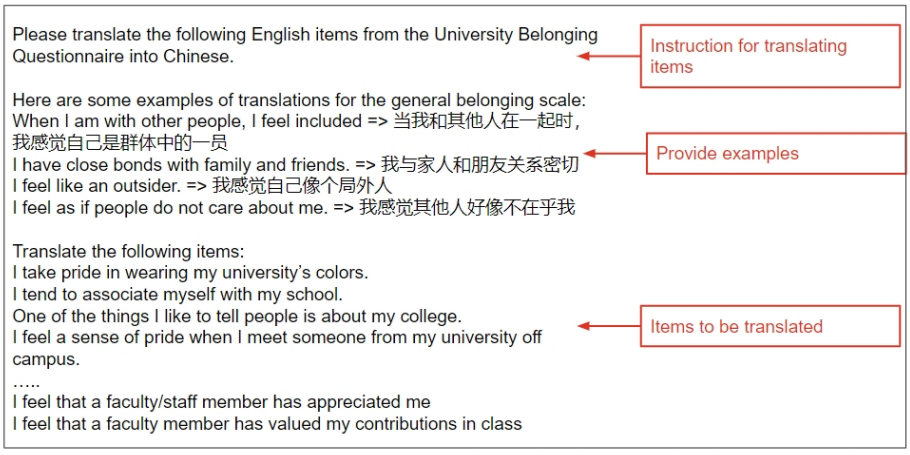

Figure 5: A item translation prompt template

shot learning, in the prompt (see Figure 5), we provided GPT with example translations from a similar scale, the General Belongingness Scale (Malone et al., 2012), which had been translated by human experts and validated in the Chinese context (Hu and Tan, 2024).

3.2. Expert Review

A group of experts with experience in educational assessment and cross-cultural studies in both the U.S. and China were invited to review the items. They first evaluated whether the original 24 items fit the Chinese context, resulting in the removal of four items (A01, A06, A08, and F04). For example, item A01, "I take pride in wearing my university's colors," was removed because most Chinese universities do not have official colors.

Next, the experts evaluated the newly generated items. For the existing dimensions, they selected the top three items from the five generated by GPT. For the new peer relationship dimension, they chose five items from the 10 generated. This ensured that each dimension would contain at least five items for subsequent psychometric analysis. The experts also reviewed the quality of the translations and provided minor suggestions for improvement, which were implemented while retaining the majority of the GPT-generated content.

After the expert review, 20 original items and 14 new items were retained, resulting in a final scale of 34 items in Chinese.(see **Appendix B**).

3.3. Psychometric Analysis

We followed the *Standards for Educational and Psychological Testing* (American Educational Research Association and on Measurement in Education, 2014) to gather evidence for the psychometric properties of the revised UBQ. This includes reliability, validity evidence related to content, internal structure, and external variables. Content validity was ensured during the expert review, while the remaining evidence was gathered through psychometric analysis using respondent data.

Data Collection. We collected responses from 318 college students in China using the expanded and translated UBQ. To assess external validity, we also administered the General Belongingness Scale (Malone et al., 2012), given its conceptual similarity to the UBQ.

Psychometric Model. The Partial Credit Model (PCM), a polytomous extension of the Rasch model, was used to assess the psychometric properties of the expanded and translated UBQ. As shown in Equation (1), the PCM describes the probability of a categorical response as a function of the respondent's trait level (i.e., the four dimensions of UBQ in this study) and the difficulty of the item. Given that the UBQ contains four dimensions, we employed a four-dimensional PCM.

$$P(y_{pi} = x|\theta_{pd}, \boldsymbol{\delta_i}) = \frac{exp[\Sigma_{k=0}^{x}(\theta_{pd} - \delta_{ix})]}{\Sigma_{h=0}^{m_i}[exp\Sigma_{k=0}^{h}(\theta_{pd} - \delta_{ik})]}, x = 0, 1, ..., m_i \tag{1}$$

where p, i, x, and d index respondent, item, category(i.e., the category respondents choose for the Likert scale item), and dimension, respectively; and y, θ, and δ denote observed responses, respondents' latent traits, and item-by-category difficulty, respectively; m_i represents the maximum scores of item i.

Reliability was quantified using both the Expected A Priori (EAP) reliability coefficient from the PCM and Cronbach's α. A Cronbach's α value of 0.70 or higher is generally considered acceptable for low-stakes assessments (Tavakol and Dennick, 2011).

The internal structure was assessed using the correlation coefficients among the UBQ dimensions, item fit statistics, and a Wright map. Correlation coefficients were derived from the multidimensional PCM, and we expected the new dimensions to show similar correlations to the existing ones. Item fit was evaluated using information-weighted fit (infit) statistics, which measure the degree of deviation between observed and expected scores. An infit of 1.0 is ideal, with values between 0.5 and 1.5 being acceptable. Values below 0.5 indicate artificially inflated reliability, while values above 1.5 suggest the item distorts measurement (Linacre, 2002, 2011). At the measurement level, the Wright map visually aligns respondents' latent trait estimates with item thresholds, with well-spaced cut points indicating a strong internal structure.

For external validity, we calculated Pearson correlation coefficients between the latent trait estimates from the UBQ and the General Belongingness Scale. Given the conceptual similarities and differences in context, we anticipated moderate correlations between the two measures. The General Belongingness Scale consists of 12 items on a 7-point Likert scale. We calculated the scores for each dimension and total scores after reversing the rejection items.

4. Results

4.1. Reliability

As shown in Table 1, the expanded and translated UBQ, generated with the assistance of GPT, demonstrated satisfactory reliability across the entire scale as well as within each individual dimension. While the reliability of the newly added dimension(0.90), *Peer Relationship*, is slightly lower than that of the original dimensions, it still meets acceptable standards. Moreover, the GPT expanded and translated scale, improved the reliability of the original scale($\alpha = 0.94$, adjusted $\alpha = 0.96$, while corrected by Spearman-Brown formula) to 0.98.

Table 1: Reliability of the new UBQ

Dimensions	Cronbach's α	EAP reliability
UBQ	0.98	0.96
University affiliation	0.94	0.91
University support and acceptance	0.94	0.88
Faculty and staff relations	0.91	0.83
Peer relationship	0.90	0.80

4.2. Evidence Related to Internal Structure

Correlation among Dimensions. The correlation matrix presented in Table 2 shows moderate correlations(i.e., 0.81-0.87) between the four dimensions, reflecting both shared

characteristics in measuring university belonging and the specific focus of each dimension. This indicates a reasonable internal structure, with the dimensions complementing one another while assessing distinct aspects of belonging.

Table 2: Correlation coefficients among the dimensions within the UBQ expanded and translated by GPT

affiliation and acceptance relations relationship	University University support Faculty and staff Peer			
University affiliation	1			
University support and acceptance	0.86	1		
Faculty and staff relations	0.87	0.86	1	
Peer relationship	0.87	0.85	0.81	1

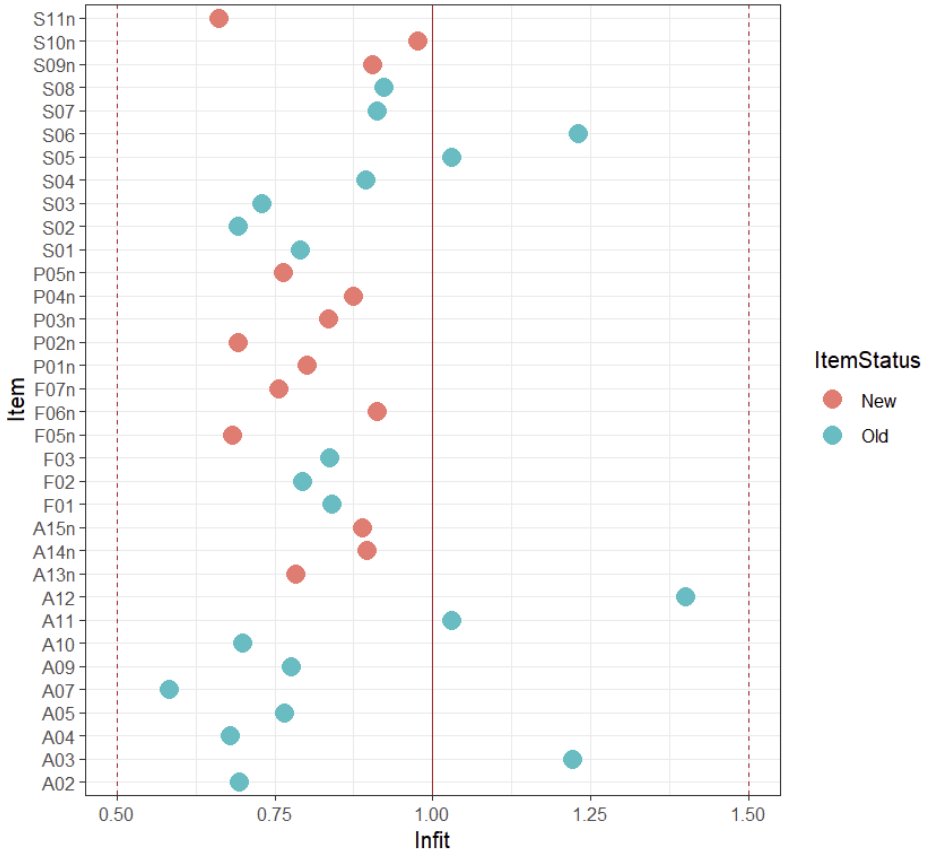

Figure 6: Infit statistics for the 34 UBQ items

Infit. The item fit statistics, presented in Figure 6, show that all 34 UBQ items fall within the acceptable range of infit values (0.5 to 1.5), indicating good item fit overall. However, it is noteworthy that the infit statistics for many items, particularly the new items generated by GPT, are generally lower than 1.0. This suggests that while GPT can effectively generate and translate UBQ items, the model tends to introduce some measurement noise, leading to slightly inflated internal consistency.

Wright Map. The Wright map, shown in Figure 7, provides a visual representation of the item distribution and latent trait levels of the respondents. Clear cut points are evident for most items, supporting the internal structure of the scale. However, two items, A09 from the original UBQ item pool and P05n generated by GPT, deviate from the expected pattern, suggesting potential issues in their measurement alignment. Overall, the Wright map demonstrates a satisfactory internal structure, even for the items generated by GPT.

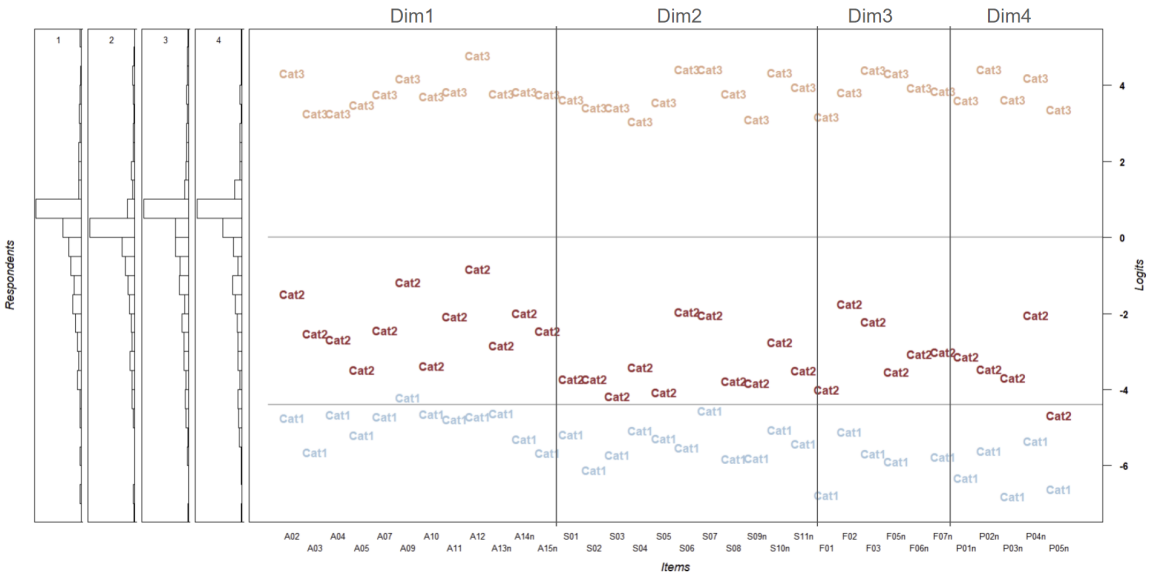

Figure 7: The Wright map
Dim1 : University affiliation; Dim2 : University support and acceptance;
Dim3 : Faculty and staff relations; Dim4 : Peer relationship

4.3. Evidence Related to External Variables.

As shown in Table 3, the four dimensions of the UBQ exhibit moderate correlations with the acceptance dimension of the General Belongingness Scale, indicating a meaningful relationship between the two constructs. Interestingly, there is also a slight, yet notable, correlation between the UBQ dimensions and the rejection dimension of the General Belongingness Scale. This result was unexpected but could be explained by the overlap in the social-emotional skills required to navigate both belonging and rejection experiences in university settings (Allen et al., 2021). In other words, students with strong social-emotional skills are likely to be sensitive to both feelings of belonging and rejection in their academic environment.

The UBQ dimensions also show a weak association with the total score of the General Belongingness Scale, though this correlation(i.e., range between 0.27 and 0.34) is lower than anticipated. This might suggest that while both scales assess related constructs, the contexts in which they operate (general versus university-specific belonging) may account for the weaker overall correlation.

Table 3: Correlation coefficents between the UBQ and the General Belongingness Scale

affiliation and acceptance relations relationship	University University support Faculty and staff Peer			
Acceptance	0.74	0.76	0.70	0.77
Rejection	0.15	0.12	0.14	0.09
General Belongingness	0.28	0.32	0.27	0.34

5. Conclusions

Using the UBQ as a case study, we demonstrated GPT's potential in expanding and translating items for non-cognitive assessments, including the introduction of new dimensions. Psychometric analyses confirmed the sound properties of these items, showcasing the model's utility in addressing two critical NLP tasks: item development and translation in non-cognitive assessments. This is particularly significant as the demand for diverse, tailored assessments continues to grow. Finally, it is important to note that item generation and translation are the first few steps of the development of non-cognitive assessments, and field tests and psychometric analyses are needed to guarantee the qualities before broader usage.

However, while the expansion and translation of the UBQ were driven by practical needs, this study is limited by focusing on a single example. Future research could extend this work by applying GPT across a broader range of non-cognitive educational assessments. Additionally, incorporating techniques such as chain-of-thought prompting and comparing GPT's performance with other large language models would provide valuable insights for practitioners in the field of educational assessment.

References

Mi Young Ahn and Howard H Davis. Students' sense of belonging and their socio-economic status in higher education: a quantitative approach. *Teaching in Higher Education*, 28 (1):136–149, 2023.

Kelly-Ann Allen, Margaret L Kern, Christopher S Rozek, Dennis M McInerney, and George M Slavich. Belonging: A review of conceptual issues, an integrative framework, and directions for future research. *Australian Journal of Psychology*, 73(1):87–102, 2021.

Kelly-Ann Allen, DeLeon L Gray, Roy F Baumeister, and Mark R Leary. The need to belong: A deep dive into the origins, implications, and future of a foundational construct. *Educational Psychology Review*, 34(2):1133–1156, 2022.

American Psychological Association American Educational Research Association and National Council on Measurement in Education. *Standards for Educational and Psychological Testing*. American Educational Research Association, 2014. doi: https://doi.org/https://doi.org/10.1787/01820d6d-en.

Yigal Attali, Andrew Runge, Geoffrey T LaFlair, Kevin Yancey, Sarah Goodwin, Yena Park, and Alina A Von Davier. The interactive reading task: Transformer-based automatic item generation. *Frontiers in Artificial Intelligence*, 5:903077, 2022.

Shreya Bhandari, Yunting Liu, Yerin Kwak, and Zachary A Pardos. Evaluating the psychometric properties of chatgpt-generated questions. *Computers and Education: Artificial Intelligence*, page 100284, 2024.

Richard W Brislin. Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology*, 1(3):185–216, 1970.

Tom B Brown. Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*, 2020.

Pedro Carneiro, Claire Crawford, and Alissa Goodman. The impact of early cognitive and non-cognitive skills on later outcomes. *CEE Discussion Papers*, 2007.

Jacob Devlin. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

Susan P Farruggia, Cheon-woo Han, Lakeshia Watson, Thomas P Moss, and Bette L Bottoms. Noncognitive factors and college student success. *Journal of College Student Retention: Research, Theory & Practice*, 20(3):308–327, 2018.

Mark J Gierl and Thomas M Haladyna. *Automatic item generation: Theory and practice*. Routledge, 2013.

Annie Gowing. Peer-peer relationships: A key factor in enhancing school connectedness and belonging. *Educational and Child Psychology*, 36(2):64–77, 2019.

Wenyan Hu and Shilin Tan. Exploring the influencing mechanisms of neighborhood environment on the physical and mental health of chinese migrant children. *The Journal of Early Adolescence*, page 02724316241261403, 2024.

John Eric Humphries and Fabian Kosse. On the interpretation of non-cognitive skills–what is being measured and why it matters. *Journal of Economic Behavior & Organization*, 136:174–185, 2017.

Aaron J Kaat, Chester " Trip" Buckenmaier, Karon F Cook, Nan E Rothrock, Benjamin D Schalet, Richard C Gershon, and Mark S Vrahas. The expansion and validation of a new upper extremity item bank for the patient-reported outcomes measurement information system®(promis). *Journal of Patient-Reported Outcomes*, 3:1–8, 2019.

Tim Kautz, James J Heckman, Ron Diris, Bas Ter Weel, and Lex Borghans. Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success. *NBER Working Paper*, 2014.

Ghader Kurdi, Jared Leo, Bijan Parsia, Uli Sattler, and Salam Al-Emari. A systematic review of automatic question generation for educational purposes. *International Journal of Artificial Intelligence in Education*, 30:121–204, 2020.

Suzanne Lane, Mark R Raymond, Thomas M Haladyna, et al. *Handbook of test development*, volume 2. Routledge New York, NY, 2016.

Antonio Laverghetta Jr and John Licato. Generating better items for cognitive assessments using large language models. In *Proceedings of the 18th workshop on innovative use of NLP for building educational applications (BEA 2023)*, pages 414–428, 2023.

Philseok Lee, Shea Fyffe, Mina Son, Zihao Jia, and Ziyu Yao. A paradigm shift from "human writing" to "machine generation" in personality test development: An application of state-of-the-art natural language processing. *Journal of Business and Psychology*, 38(1): 163–190, 2023.

M Linacre. What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions*, 16:878, 2002.

M Linacre. Teaching rasch measurement. *Rasch Measurement Transactions*, 31:1630, 2011.

Glenn P Malone, David R Pillow, and Augustine Osman. The general belongingness scale (gbs): Assessing achieved belongingness. *Personality and Individual Differences*, 52(3): 311–316, 2012.

OECD. *PISA 2022 Technical Report*. OECD Publishing, 2024. doi: https://doi.org/https://doi.org/10.1787/01820d6d-en. URL https://www.oecd-ilibrary.org/content/publication/01820d6d-en.

Uldis Ozolins, Sandra Hale, Xiang Cheng, Amelia Hyatt, and Penelope Schofield. Translation and back-translation methodology in health research–a critique. *Expert Review of Pharmacoeconomics & Outcomes Research*, 20(1):69–77, 2020.

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. Language models are unsupervised multitask learners. *OpenAI Blog*, 1(8):9, 2019.

Christopher D Slaten, Zachary M Elison, Eric D Deemer, Hayley A Hughes, and Daniel A Shemwell. The development and validation of the university belonging questionnaire. *The Journal of Experimental Education*, 86(4):633–651, 2018.

Mohsen Tavakol and Reg Dennick. Making sense of cronbach's alpha. *International Journal of Medical Education*, 2:53, 2011.

Luis Francisco Vargas-Madriz and Chiaki Konishi. The relationship between social support and student academic involvement: The mediating role of school belonging. *Canadian Journal of School Psychology*, 36(4):290–303, 2021.

Mark Wilson. *Constructing measures: An item response modeling approach*. Routledge, 2023.

Qin Xie, Xiaoling Zhong, Wen-Chung Wang, and Cher Ping Lim. Development of an item bank for assessing generic competences in a higher-education institute: a rasch modelling approach. *Higher Education Research & Development*, 33(4):821–835, 2014.

Xuchen Yao, Gosse Bouma, and Yi Zhang. Semantics-based question generation and implementation. *Dialogue and Discourse*, 3(2):11–42, 2012.

Appendix A. 24 Items of the Original UBQ

As shown in Figure 8, there are 24 items in the original UBQ. 12 items assess the university affiliation; eight items assess university support and acceptance; four items are about faculty and staff relations.

No.	Dimension	Item
A01	University affiliation	I take pride in wearing my university's colors
A02		I tend to associate myself with my school
A03		One of the things I like to tell people is about my college
A04		I feel a sense of pride when I meet someone from my university off campus
A05		I would be proud to support my university in any way I can in the future
A06		I have university-branded material that others can see
A07		I am proud to be a student at my university
A08		I attend university sporting events to support my university
A09		I feel "at home" on campus
A10		I feel like I belong to my university when I represent my school off campus
A11		I have found it easy to establish relationships at my university
A12		I feel similar to other people in my major
S01	University support and acceptance	My university provides opportunities to engage in meaningful activities
S02		I believe there are supportive resources available to me on campus
S03		My university environment provides me an opportunity to grow
S04		My university provides opportunities to have diverse experiences
S05		My cultural customs are accepted at my university
S06		I believe I have enough academic support to get me through college.
S07		I am satisfied with the academic opportunities at my university
S08		The university I attend values individual differences
F01	Faculty and staff relations	I believe that a faculty/staff member at my university cares about me
F02		I feel connected to a faculty/staff member at my university
F03		I feel that a faculty/staff member has appreciated me
F04		I feel that a faculty member has valued my contributions in class

Figure 8: The Original UBQ items in English

Appendix B. 34 Items of the UBQ expanded and translated by GPT

As shown in Figure 9, after item generation and expert review, A01, A06, A08, and F04 were removed to better accommodate the Chinese context. Item names ending with n indicate the new ones generated by GPT. Chinese translation by GPT is listed in the last column.

No.	Item	Translation
A02	I tend to associate myself with my school	我习惯于把自己与学校联系在一起
A03	One of the things I like to tell people is about my college	我喜欢向人分享我大学的事情
A04	I feel a sense of pride when I meet someone from my university off campus	当我在校园以外遇见来自我大学的人时，我感到自豪
A05	I would be proud to support my university in any way I can in the future	我将会非常荣幸在未来的日子里用我可以提供的任何方式来支持我的大学
A07	I am proud to be a student at my university	我为自己是这所大学的一员而感到荣耀
A09	I feel "at home" on campus	我在校园里感觉如同在家一般
A10	I feel like I belong to my university when I represent my school off campus	当我代表学校在校外活动时，我感到自己就是学校的一部分
A11	I have found it easy to establish relationships at my university	在我的大学，我发现建立关系很容易
A12	I feel similar to other people in my major	我觉得我和同专业的其他人很相似
A13n	I enjoy participating in university traditions and events.	我喜欢参加大学的传统和活动
A14n	I identify strongly as a student of my university.	我与我所在的大学建立了深厚的认同感
A15n	My university's achievements significantly add to my personal pride.	我为学校的成就而感到自豪
S01	My university provides opportunities to engage in meaningful activities	我的大学提供了参与有意义活动的机会
S02	I believe there are supportive resources available to me on campus	我相信校园内有很多可供我利用的资源
S03	My university environment provides me an opportunity to grow	我的大学环境给我提供了成长的机会
S04	My university provides opportunities to have diverse experiences	我的大学提供了丰富体验的机会
S05	My cultural customs are accepted at my university	我大学接受我的文化习俗
S06	I believe I have enough academic support to get me through college.	我相信我拥有足够的学术支持来帮助我完成大学学业
S07	I am satisfied with the academic opportunities at my university	我对大学的学术机会感到满意
S08	The university I attend values individual differences	我所在的大学重视个体差异
S09n	My university encourages my participation in activities and initiatives.	我的学校鼓励我积极参与各种活动和计划
S10n	I feel my personal needs and wellness are prioritized by my university.	我觉得学校把我的需要和健康放在了首位
S11n	I think my university gives me the necessary platform to express my views and opinions freely.	我觉得我的学校为我提供了一个自我表达的舞台
F01	I believe that a faculty/staff member at my university cares about me	我相信在学校有老师会关心我
F02	I feel connected to a faculty/staff member at my university	我觉得我与学校的某位教职员工建立了联系
F03	I feel that a faculty/staff member has appreciated me	我觉得大学里有教职工欣赏我
F05n	I have found that staff and faculty members are approachable and friendly.	我发现学校的教职员工既友好又亲切
F06n	I feel that faculty/staff members understand and respect my perspective and ideas.	我觉得教职工能理解和尊重我的观点和想法
F07n	I believe faculty members invest time and effort in my academic journey.	我觉得教职员工为我的学术成长投入了时间和努力
P01	I feel comfortable and accepted around my classmates in the university.	在大学同学中，我感到舒适和被接纳
P02	My fellow students include me in social activities and discussions.	我的同学在社交活动和讨论中都让我参与其中
P03	I have trustworthy friends at the university who support me.	在大学里，我有可以信任的并且会支持我的朋友
P04	I feel there is a strong sense of community among my peers in the university.	我在大学里感到了强烈的凝聚力
P05	My peers help me if I face any difficulties with my academics or university life.	如果在学业中遇到困难，我的大学同伴会帮助我

Figure 9: UBQ items expanded and translated by GPT

Appendix C. Survey instruction

亲爱的同学:

您好！我们是 xx 大学 xx 部门的 xx 研究人员，现邀请您参加 “xx 研究” 。

在您决定是否参加这项研究之前，请尽可能详细地阅读以下内容。如有疑问，您可以请研究团队给予解释，帮助您更好地做出决定。

一、项目简介：

本项目旨在了解教师和学生的日常工作、学习和生活状况。希望能够通过问卷调查，了解教师和学生的状态和面临的问题，并为他们状态的改变和问题的解决提供建议。希望得到您的理解和支持。

二、研究过程：

本次调研将从各个角度了解您的状态，包括基本情况、您的情绪、感受、工作状态等。填写的时间大概 20-30 分钟。

三、参加研究可能的风险：

本研究调查的内容不会造成伤害。如果您对问卷中涉及的任何问题觉得不舒服，可以随时告诉我们，停止该部分问题，直接进入下一部分。您也可以随时提出停止调查，我们将尊重您的意愿。您的权益不会因此受到影响。

四、参加研究可能的受益：

您提供的信息将有益于高校教师和学生的研究工作。

五、保密性：

您在调查中提供的个人资料均属保密。您的测量记录将完整地保存在研究中，仅用于科学研究。任何有关本研究结果的公开报告都不会披露您的个人信息。

六、自愿原则：

您可以随时了解本研究相关信息，自愿决定是否继续参与。参与过程中，如果感到任何不适，您可以通知研究团队或要求退出。您的权益不会因此受到影响。

Figure 10: Survey instruction in Chinese

Dear Student,

Hello! We are researchers from the [Department] of [University Name], and we are inviting you to participate in the "[Research Study Name]" project.

Before deciding whether to participate, please read the following information as carefully as possible. If you have any questions, feel free to ask the research team for clarification to help you make a better-informed decision.

1. Project Overview:

This project aims to understand the daily work, study, and life conditions of teachers and students. Through a questionnaire survey, we hope to gain insights into the challenges and experiences of both groups and provide recommendations for improving their conditions and addressing their issues. We would appreciate your understanding and support.

2. Research Process:

The survey will gather information about your overall status from various aspects, including basic information, emotions, feelings, and work conditions. It will take approximately 20-30 minutes to complete.

3. Potential Risks of Participation:

The content of this survey poses no risk of harm. If any questions in the questionnaire make you uncomfortable, you can inform us at any time, skip those questions, and proceed to the next section. You may also choose to withdraw from the survey at any time, and we will respect your decision. Your rights will not be affected.

4. Potential Benefits of Participation:

The information you provide will contribute to research focused on improving the well-being of university teachers and students.

5. Confidentiality:

Any personal information you provide in the survey will remain confidential. Your responses will be securely stored and only used for scientific research. Public reports based on the research findings will not disclose any personal information.

6. Voluntary Participation:

You are free to inquire about the research at any time and voluntarily decide whether to continue participating. If you experience any discomfort during the study, you may notify the research team or choose to withdraw. Your rights will not be affected by this decision.

Figure 11: Survey instruction in Enligsh