Development of a College Student Validation Survey: A Design-Based Research Approach

Toni A. May Dara N. Bright Yiyun (Kate) Fan Christopher Fornaro
Kristin L. K. Koskey Thomas Heverin

Rendón's (1994) seminal research on validation theory (VT) provided a model for understanding how validating experiences can positively influence "culturally diverse" (p. 33) students in higher education. Validation is "an enabling, confirming and supportive process initiated by in- and out-of-class agents that fosters academic and interpersonal development" (Rendón, 1994, p. 44) and is critical for the transition, persistence, and success of college students (Rendón, 1994, 2002). Through this theoretical model, scholars have extensively explored how institutions can provide validating experiences by developing supportive learning environments for general undergraduate populations and specific groups such as Black, Latinx, low-income, first-generation, and two-year college students (e.g., Allen, 2016; Bauer, 2014). Many prior studies have relied on qualitative methods. While Rendón and Muñoz (2011) have called for further study of validation's impact on student outcomes through quantitative methods, few quantitative instruments of VT exist. The primary tool used for assessing VT consists of two scales from the larger Diverse Learning Environments (DLE; Hurtado et al., 2011) survey that have demonstrated their effectiveness for measuring academic validation in class and general interpersonal validation among college students at large (Hurtado et al., 2015). DLE scales were not, however, designed to match Rendón's full four-component conception of VT (i.e., academic in-class, academic out-of-class, interpersonal in-class, interpersonal out-of-class). Thus, a new measure of VT is necessary to capture quantitative information aligned with Rendón's model. The purpose of this study was to expand the field of quantitative VT research by presenting validity evidence from a new survey entitled the Validation Theory Survey (VTS) that was designed to align with Rendón's VT model and to be used with undergraduate students. One overarching research question guided this study: To what extent did validity evidence (i.e., content, response process, consequential, and internal structure) support the use of the VTS to evaluate undergraduates' perceptions of their academic and interpersonal validating experiences inside and outside higher education classrooms?

Toni A. May is Associate Professor of Assessment, Research, and Statistics; Dara N. Bright, Yiyun (Kate) Fan, and Christopher Fornaro are doctoral candidates in the School of Education; Kristin L. K. Koskey is Research Professor in the School of Education; and Thomas Heverin is Associate Teaching Professor of Computing and Security Technology in the College of Computing and Informatics; all at Drexel University.

METHODS

Educational design-based research (DBR) approaches emphasize a process of developing tools for a specific purpose through iterative methods of designing, testing, evaluating, and reflecting (Scott et al., 2020). DBR techniques implemented to develop and validate educational instruments have been effective when engaging in qualitative and quantitative field-testing methods (e.g., Sondergeld & Johnson, 2019) to evaluate multiple sources of validity evidence in concordance with *The Standards for Educational and Psychological Testing* (AERA et al., 2014). *The Standards* have urged instrument developers to evaluate multiple types of validity evidence, including (a) content or item alignment with construct; (b) response process, that is, participants understand the instrument as researchers intended; (c) consequential or controlling bias or potential negative impact on participants; (d) internal structure, that is unidimensional and reliable constructs formed; and (e) relationship to other variables or ensuring instrument outcomes are related to other hypothesized variables. This study presents findings related to the development and validation of the four VTS scales and reports on all types of validity evidence except the relationship to other variables, which will be examined in subsequent research.

Instrumentation

As part of a grant managed by the National Science Foundation in collaboration with the U.S. Office of Personnel Management and Department of Homeland Security, the VTS was created to evaluate project impact on undergraduate cybersecurity majors' validating experiences in the program compared to those not participating. After a thorough review and synthesis of VT literature (e.g., Acevedo-Gil et al., 2015; Allen, 2016; Baber, 2018; Bauer, 2014; Rendón, 1994, 2002; Rendón Linares & Muñoz, 2011), 44 unique items were drafted to represent commonly noted validating experiences (17 academic in-class; 9 academic out-of-class; 9 interpersonal in-class; 9 interpersonal out-of-class). These were rated on a 5-point scale (*strongly disagree, disagree, agree, strongly agree, not applicable*). Survey refinement based on two rounds of qualitative- and quantitative-field testing resulted in a final survey comprised of 26 items (7 academic in-class; 6 academic out-of-class; 7 interpersonal in-class; 6 interpersonal out-of-class) rated on a 4-point scale (*not applicable* removed).

DATA COLLECTION AND ANALYSIS

Qualitative Field-Testing

An expert panel (n = 9) in relevant content domains (i.e., higher education; student development; diversity, equity, and inclusion; psychometrics) evaluated original and modified VTS items through questionnaires to indicate item-to-theory alignment and assessed content validity through their lens of expertise. Cognitive interviews with a diverse group of undergraduates (n = 13) were conducted to assess response process validity. Two strategies were implemented to help students feel comfortable providing critical feedback on the VTS. First, three trained graduate students conducted the interviews as near-peers. Second, the interviewers' stated purpose was to get feedback on the "clarity and understanding" of the items. As such, they told students, "Don't worry about hurting our feelings if you criticize the questions. Our job is to find out if they are functioning well and how we can make them better." Responses were first coded for alignment with item intention (aligned/not aligned) and second for emergent themes. At the completion of the cognitive interviews, undergraduates were asked three questions to determine whether

completing the survey caused them to feel any negative reactions, which investigated consequential validity. Qualitative data were analyzed through conceptual content analysis (Christie, 2007) to allow for the identification of concepts in the text and their frequency.

Quantitative Field-Testing

Once VTS items were modified from qualitative findings, the survey underwent two quantitative field-testing rounds. The survey sample for this study consisted of 215 undergraduates taking courses in a College of Computing and Informatics from a 4-year private R1 university located in the Northeast. Students were sent an electronic survey link. Slightly more than half of the participants (57%) identified as male. Approximately one third identified as either Asian (39%) or White (31%). Nearly two thirds of survey participants indicated they were at the end of their first year (63%). Rasch (1960/1980) measurement was implemented using Winsteps (Linacre, 2021) to examine the four VTS constructs with the Rasch rating scale model (Andrich, 1978). Rasch measurement was used because of its effectiveness in survey development and demonstrated advantages over classical test theory methods (see Bond & Fox, 2007). While a comprehensive review is not possible in this manuscript, two strengths include sample independence within standard error bands and the model's ability to handle missing data due to its probabilistic nature. Linacre's (2002) guidelines for optimizing rating scale categories allowed for the quantitative study of response process validity. A variety of psychometric indices were used to evaluate the internal structure validity of the VTS, including Rasch item fit statistics (i.e., infit, outfit, point-biserial), Rasch Principal Components Analysis (RPCA), item and person reliability and separation indices, and a variable map (Wright map) to evaluate item measure redundancy.

RESULTS

Content Validity Evidence

Qualitative field-testing findings showed all expert panelists reported 100% of VTS items aligned with their intended validation theory construct. Two main themes identified from expert feedback—word choice and new items needed—resulted in 10 item revisions. Most item modifications were related to word choice. For instance, instead of asking if instructors expressed "concern" for teaching students, experts suggested a shift toward a different word, such as "interest." Another item revision example from expert feedback was to break a broad item related to student educational pursuit after college down into two separate items focusing on professional aspirations and career paths to better capture undergraduate short- and long-term goals.

Response Process

Nearly all (91.2%) undergraduate cognitive interview responses aligned with the researchers' intended item meaning. During cognitive interviews, undergraduates provided feedback for item revision in three ways: (a) by clarifying prompts, (b) by suggesting where examples were needed, and (c) by questioning response options. In an attempt to improve item clarity, the phrase "When I am in class" or "When I am out of class" was added to every question. Initially, this framing was provided at the beginning of the item set rather than at the start of each item. Some undergraduates responded with examples that did not coincide with the intended VT domain. For example, one undergraduate shared a misaligned response of how their previous academic experiences helped during job interviews while responding to an academic-in-class item. Providing examples within some items was recommended by undergraduates to help them better respond to the intended item's meaning. For example, when undergraduates

Table 1.
Summary of Final Quantitative Field-Testing VTS Construct Findings

	Survey construct			
Guidelines	**Academic in-class**	**Academic out-of-class**	**Interpersonal in-class**	**Interpersonal out-of-class**
Items / Students measured	7 / 157	6 / 158	7 / 181	6 / 179
Scale function				
At least 10 observations per category	Yes	Yes	Yes	Yes
Average category measures advance monotonically	Yes	Yes	Yes	Yes
Outfit MNSQ < 2.0	Yes	Yes	Yes	Yes
Step difficulties advance by at least 1.4 logits	Yes	Yes	Yes	Yes
Reliability (< .70 = Poor; .70 = Acceptable; .80 = Good; .90 = Excellent)				
Person	0.89	0.80	0.78	0.76
Item	0.96	0.96	0.97	0.97
Separation (< 1.5 = Poor; 1.5 = Acceptable; 2.0 = Good; 3.0 = Excellent)				
Person	2.78	2.03	1.90	1.80
Item	4.70	4.76	5.82	5.79
Point-Biserial (Positive value required)				
Items with negative point-biserial	None	None	None	None
Fit (MNSQ > 2.0 = Degrades measure; < 0.5 or > 1.5 = Less productive, not degrading; 0.5 to 1.5 = Productive for measure)				
Items misfitting	None	Item 5	None	Item 3
Unidimensionality (RPCA > 50% = Good)				
RPCA	61.1%	61.0%	61.2%	56.6%

Note. Guideline citations: Scale function (Linacre, 2002); reliability and separation (Duncan et al., 2003); item fit (Linacre, 2002); RPCA (Linacre, 2022).

were asked if the curriculum reflected their personal background, some were unsure what this meant. Two undergraduates shared that personal background meant "life and academic experience," so an example drawn directly from cognitive interview feedback was added to this item. Finally, there was some confusion around "Not Applicable" as a response option. Students who had an answer when they were given a chance to talk aloud sometimes selected "Not Applicable." For example, on one item, an undergraduate shared that if it "never happened" to them, they would select "Not Applicable" or "Disagree." To eliminate confusion in response choices, the "Not Applicable" option was removed before quantitative field testing.

Quantitative field testing data also informed response process validity evidence. Alignment of Rasch findings with scale optimization guidelines (see Table 1) supported undergraduates' use of the 4-point scale on the VTS as intended. For each rating scale category: at least 10 observations were noted. All advanced monotonically, outfit mean-square was less than 2.0, and step calibrations advanced by at least 1.4 logits from one to the next.

Table 2.
Final VTS Item Stems by Construct in
Order of Difficulty Measure with SE and Item Statistics

VTS construct item stem	Item statistics			
	Measure in logits (SE)	Infit (MNSQ)	Outfit (MNSQ)	Point-biserial
Academic in-class (7 items)				
My instructors are approachable for academic questions.	–2.07 (0.23)	0.64	0.59	0.82
My instructors structure learning experiences that make me capable of learning.	–1.15 (0.23)	0.80	0.63	0.83
The classroom environment encourages me to participate in course activities.	–0.57 (0.48)	0.85	0.75	0.82
My instructors discuss possible career paths in classroom activities.	–0.16 (0.22)	1.15	1.05	0.79
My instructors work individually with me when I need help.	0.79 (0.21)	1.20	1.05	0.82
My instructors consider my academic interests in classroom activities.	1.33 (0.20)	0.96	0.93	0.82
My instructors recognize my academic improvement.	1.85 (0.19)	0.93	1.08	0.85
Academic out-of-class (6 items)				
My friends/family encourage me to share academic successes with them.	–2.57 (0.54)	0.77	0.77	0.81
My friends/family encourage me to discuss academic challenges with them.	–1.02 (0.48)	1.75	1.45	0.79
My classmates and I give each other academic support.	–0.89 (0.25)	0.60	0.37	0.86
My counselor/advisor provides me with academic advice.	1.20 (0.22)	0.97	0.63	0.82
My instructor provides me with opportunities to reflect on my learning.	1.54 (0.22)	1.18	0.78	0.76
My counselor/advisor helped me develop my academic plan of study.	1.73 (0.22)	0.86	0.64	0.84
Interpersonal in-class (7 items)				
My instructors refer to me by my preferred name.	–2.59 (0.24)	0.64	0.68	0.85
My instructors correctly pronounce my name.	–1.86 (0.25)	0.80	0.71	0.88
My instructors treat all students fairly.	–1.06 (0.44)	1.01	0.72	0.73
My identity is accepted by my instructor.	–1.03 (0.44)	0.94	0.69	0.74
My instructors encourage me to interact with other students.	0.91 (0.21)	1.03	1.40	0.81
My instructors see my previous personal life experiences as contributing to the learning environment.	2.60 (0.38)	1.42	1.50	0.69

Table 2, continued.

VTS construct item stem	Item statistics			
	Measure in logits (SE)	Infit (MNSQ)	Outfit (MNSQ)	Point-biserial
My instructors see my personal identity as adding value to the class.	3.02 (0.37)	1.10	1.32	0.79
Interpersonal out-of-class (6 items)				
My identity is accepted in college activities/ organizations.	–2.57 (0.40)	1.46	0.67	0.83
My friends/family provide me with personal encouragement.	–1.62 (0.20)	0.89	1.00	0.80
Student organizations I participate in value my contributions.	–0.98 (0.44)	1.94	0.69	0.77
Campus sponsored activities encourage me to socially engage with other students.	1.34 (0.33)	0.76	1.82	0.63
My advisor/counselor has suggested resources related to my personal needs.	1.75 (0.18)	0.99	0.95	0.68
My college sponsors events for families and students to attend together.	2.09 (0.17)	0.72	1.52	0.46

Note. Mean item difficulty is set at 0 logits.

Consequential

All undergraduates willingly completed the survey during their cognitive interview and did not skip any questions asked of them even though they knew participation was voluntary and they could stop at any time for any reason. No undergraduates mentioned perceiving any items to be biased. One student noted some items could potentially "trigger a negative response" if a student had a previous negative experience they were reflecting on while answering. This feedback prompted the research team to add language and contact information for the university counseling center at the end of the survey in the event students wanted to speak with someone further.

Internal Structure

Table 1 shows a summary of the final psychometric properties from iterative polytomous Rasch analyses of the VTS constructs. A total of 18 original VTS items were removed from across constructs either due to misfit or redundancy in item measure difficulty. To summarize psychometric results across the final survey constructs: reliability (person range = 0.76–0.89; item range = 0.96–0.97) and separation (person range = 1.80–2.78; item range = 4.70–5.82) were acceptable or better. No items had a negative point-biserial, no final items misfit to the point of degrading the measure (MNSQ < 2.0), and RPCA was greater than 50% across constructs. Collectively, these results suggest unidimensional constructs were formed by VTS items in each of the four scales. See Table 2 for final VTS items representing examples of validating experiences by construct with item statistics.

DISCUSSION

This study adds to the literature by presenting four validated survey scales aligned with Rendón's (1994, 2002) vision of VT for college

undergraduates and further expands the field of study related to college students' academic and interpersonal validating experiences in and out of the classroom. Multiple aspects of validity evidence collected from varying qualitative and quantitative data sources through a robust validation study align with educational survey development DBR methodologies (Scott et al., 2020) and *The Standards* (AERA et al., 2014). Results from this study provided empirical evidence, which led to the removal of misfitting and redundant items, resulting in more parsimonious VTS scales. We were able to shorten the overall length and the time required for completion while maintaining strong psychometric properties. It is not recommended to form an overall score using all 26 items. Composite scores for each of the four scales should be formed since each scale functioned as a unidimensional construct. Examining each component separately as part of internal (self-assessment) or external evaluation allows for guiding conversations around a specific program or institutional strengths and areas for improvement in validating experiences for undergraduates. While our findings support using the VTS with undergraduates, the sample was limited to computing and informatics undergraduates at a private R1 university who primarily identified as Asian or White. Thus, additional VTS research is needed with more diverse undergraduates and in other university contexts.

Correspondence concerning this article should be addressed to Toni A. May at Drexel University; tas365@drexel.edu

REFERENCES

Acevedo-Gil, N., Santos, R. E., Alonso, Ll., & Solorzano, D. G. (2015). Latinas/os in community college developmental education: Increasing moments of academic and interpersonal validation. *Journal of Hispanic Higher Education, 14*(2), 101–127. https://doi.org/10.1177/1538192715572893

Allen, T. O. (2016). (In)validation in the minority: The experiences of Latino students enrolled in an HBCU. *The Journal of Higher Education, 87*(4), 461–487.

American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (2014). *Standards for educational and psychological testing.* American Educational Research Association.

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43*(4), 561–573.

Baber, L. D. (2018). "Living in the along": Validating experiences among urban community college students in a college transition program. *Community College Review, 46*(3), 316–340. https://doi.org/10.1177/0091552118775813

Bauer, K. (2014). Black male community college students and faculty–student engagement: Differences in faculty validation and time status. *Journal of Progressive Policy and Practice, 2*(2), 157–164.

Bond, T., & Fox, C. (2007). *Fundamental measurement in the human sciences* (2nd ed.). Erlbaum.

Christie, C. (2007). Content analysis. In R. Baumeister & K. Vohs (Eds.), *Encyclopedia of social psychology* (p. 176). SAGE.

Duncan, P. W., Bode, R. K., Lai, S. M., & Perera, S. (2003). Rasch analysis of a new stroke-specific outcome scale: The stroke impact scale. *Archives in Physical Medicine Rehab, 84*(7), 950–963.

Hurtado, S., Cuellar, M., & Wann, C. G. (2011). Quantitative measures of students' sense of validation: Advancing the study of diverse learning environments. *Enrollment Management Journal, 5*(2), 53–71.

Hurtado, S., Ruiz Alvarado, A., & Guillermo-Wann, C. (2015). Creating inclusive environments: The mediating effect of faculty and staff validation on the relationship of discrimination/bias to students' sense of belonging. *Journal Committed to Social Change on Race and Ethnicity, 1*(1), 60–80.

Linacre, J. M. (2002). What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions, 16*(2), 878.

Linacre, J. M. (2021). WINSTEPS® (Version 4.280) [Computer Software]. Beaverton, Oregon.

Linacre, J. M. (2022). *A user's guide to WINSTEPS* (Program Manual 5.2.3). https://www.winsteps.com/winman/principalcomponents.htm

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests.* (Copenhagen, Danish Institute for Educational Research), with foreword and afterword by B. D. Wright. The University of Chicago Press. (Original work published 1960)

Rendón, L. I. (1994). Validating culturally diverse students: Toward a new model of learning and student development. *Innovative Higher Education, 19*(1), 35–51.

Rendón, L. I. (2002). Community college Puente: A validating model of education. *Educational Policy, 16*(4), 642–667.

Rendón Linares, L. I., & Muñoz, S. M. (2011). Revisiting validation theory: Theoretical foundations, applications, and extension. *Enrollment Management Journal, 5*(2), 12–33.

Scott, E. E., Wenderoth, M. P., & Doherty, J. H. (2020). Design-based research: A methodology to extend and enrich biology education research. *Life Sciences Education, 19*(3), 1–12.

Sondergeld, T. A., & Johnson, C. C. (2019). Development and validation of a 21st century skills assessment: Using an iterative multimethod approach. *School Science and Mathematics Journal, 119*(6), 312–326. https://doi.org/10.1111/ssm.12355