www.ijemst.net

A Rasch Analysis of the Self-determination, Purpose, Identity, and Engagement in Science (SPIRES) Survey: Instrument Validation and Recommendations

Courtney Donovan
Colorado Denver, United States

Hannah Huvard
University of Colorado Denver, United States

Angela Rexwinkle
University of Colorado Denver, United States

Robert Talbot
University of Colorado Denver, United States

To cite this article:

Donovan, C., Huvard, H., Rexwinkle, A., & Talbot, R. (2025). A Rasch analysis of the Self-determination, Purpose, Identity, and Engagement in Science (SPIRES) Survey: Instrument validation and recommendations. *International Journal of Education in Mathematics, Science, and Technology (IJEMST), 13*(4), 992-1015. https://doi.org/10.46328/ijemst.4903

International Journal of Education in Mathematics, Science and Technology

2025, Vol. 13, No. 4, 992-1015 https://doi.org/10.46328/ijemst.4903

A Rasch Analysis of the Self-determination, Purpose, Identity, and Engagement in Science (SPIRES) Survey: Instrument Validation and Recommendations

Courtney Donovan, Hannah Huvard, Angela Rexwinkle, Robert Talbot

Article Info

Article History

Received:
15 January 2025
Accepted:
18 May 2025

Keywords

Rasch
Survey validation
SPIRES

Abstract

The purpose of the present study was to perform a cross-validation of an existing measure of science students' motivational traits using the Rasch modeling approach. The validity of the Self-determination, Purpose, Identity, and Engagement in Science (SPIRES) survey was originally investigated using factor analysis, but a secondary validation of this instrument has not yet been published. This is a recommended practice when using a psychometric instrument within a new context or with a different student population. In this validity study, we took a Rasch modeling approach instead of factor analysis because, unlike factor analysis, Rasch modeling is sample-independent. The original factor analysis validation of the SPIRES suggested the survey is composed of three larger ideas or constructs, while our Rasch modeling results suggest there are four constructs. Since our Rasch analyses were sample-independent, we conclude that the SPIRES survey is a four construct survey and may be treated as such across educational contexts without further need to validate using factor analysis, which could continuously produce inconsistent results. These results provide the basis for a validity argument for researchers using the SPIRES in their work. Our work also demonstrates an advantage to using a Rasch modeling approach over factor analysis for instrument validation.

Introduction

In recent years, higher education institutions have invested heavily in increasing student achievement within undergraduate science courses in response to low retention rates and inequity across the board in science education. Fewer than 40% of students who enter college as a STEM (science, technology, engineering, and mathematics) major actually complete a STEM degree (President's Council on Science and Technology, 2012). Additionally, the percentage of underrepresented and minority students who persist to earn a STEM degree is significantly lower than the percentage of White students who persist to earn a STEM degree (National Science Foundation, National Center for Science and Engineering Statistics, 2019; Riegle-Crumb et al., 2019). Undergraduate science students' motivational traits, like engagement, identity, and sense of belonging have been shown to be major contributors to students' learning of science and their persistence in the sciences (Huvard et al., 2020; Carlone & Johnson, 2007; Glynn et al., 2011; Jack et al., 2014; Knekta et al., 2020; Skinner et al., 2017; Stets et al., 2017; Tsai et al., 2017). As such, these traits may be used as indicators of students' achievement in individual courses and the effectiveness of various interventions. However, student motivational traits are not easily measured, and many psychometric instruments that are developed to measure them are not valid across contexts nor different student populations (Komperda et al., 2018).

The purpose of the present study was to perform a cross-validation of an existing measure of science students' motivational traits using the Rasch modeling approach. The measure, the Self-determination, Purpose, Identity, and Engagement in Science (SPIRES) survey, was originally designed by Skinner et al. (2017). A secondary validity study has not yet been published on the entire SPIRES survey, which is a recommended practice when any survey is being used in a different context from the one from which was originally developed (Komperda et al., 2018) to determine how the instrument is functioning in the new population. Within different educational settings, student populations, and time points, psychometric tests have been shown to function differently (Glynn et al., 2009, 2011). Thus, many measurement methodologists argue that instrument functioning, validity, and reliability needs to be demonstrated each time a psychometric instrument is used in a new context or different student population from which it was originally developed (Kane, 2006; Komperda et al., 2018).

Psychometric instruments such as the SPIRES survey are typically validated using factor analysis, which is sample-dependent (Lomax & Schumacker, 2010). In this validity study, we used Rasch modeling instead of factor analysis because, unlike factor analysis, Rasch modeling is sample-independent. A Rasch model assumes unidimensionality and estimates the respondents' abilities to agree with (i.e. endorse) items on the instrument probabilistically as a function of the difficulty of items themselves. Romine et al. (2017) explain that "the goal of the Rasch model is not to fit the data, but rather to evaluate how data fit with what would be expected from a well-constructed measurement tool" (p. 649). Taking a Rasch modeling approach to developing a validity argument for the SPIRES survey strengthens the legitimacy of this survey as a tool to measure students' motivational factors and aids in its validity for use across contexts. We are not aware of a Rasch analysis in the literature to examine the scales within the SPIRES survey.

The Self-determination, Purpose, Identity, and Engagement in Science (SPIRES) Survey

The Self-determination, Purpose, Identity, and Engagement in Science (SPIRES) survey is an instrument that is intended to measure science students' motivational factors. Developed by Skinner et al. (2017), this survey is composed of three sets of motivational factors: (a) Self-System Processes, (b) Student Engagement, and (c) Identity as a Scientist. Each factor consists of three or four short form scales, for a total of 11 short form scales: (a) Competency, (b) Autonomy/Ownership, (c) Relatedness/Belonging, (d) Behavioral Engagement, (e) Emotional Engagement, (f) Behavioral Disaffection, (g) Emotional Disaffection, (h) Science Identity, (i) Science Career Plans, (j) Purpose in Science, and (k) Positive Relationships and Collaborations. Each of these scales are made of 4-5 questions that are evaluated on a 5-point Likert scale (1 = strongly agree to 5 = strongly disagree).

The stated purpose of the SPIRES survey is "to help create a window into undergraduates' motivational experiences in science classes, by developing a suite of conceptually focused and psychometrically sound motivational surveys for use by instructors, researchers, and interventionists" (Skinner et al., 2017, p. 6). The original SPIRES survey was constructed based on the theoretical framework of self-determination theory (SDT), which highlights human motivation as something that is intrinsic, rather than something that is acquired through socialization (Deci & Ryan, 2000; Ryan & Deci, 2017; Skinner et al., 2017). SDT suggests that humans are intrinsically motivated when their basic psychological needs are met within a given context. These psychological needs include competence, autonomy, and relatedness (Deci & Ryan, 2000). Within a learning environment, these psychological needs have been hypothesized to be essential pieces of students' intrinsic motivation for learning and engagement. Skinner et al. (2017) call the fulfillment of these needs the "motivational fire" (p. 3) of learning and persistence in the sciences. Thus, in a science classroom, SDT argues that students must feel competent (i.e., they feel capable of studying and participating in science), autonomous (i.e., they feel that they are their one independent source of motivation), and related (i.e., they feel connected to others within the science learning context, which includes a strong sense of belonging to the science learning community).

If these basic needs of competence, autonomy, and relatedness are met, students will be more intrinsically motivated to learn and engage with science (Skinner et al., 2017). Skinner et al. (2017) refer to these needs collectively as "self-systems processes", as they pertain to students' own views of themselves within the learning environment (p. 5). Skinner et al. (2017)'s conceptualization of the learning of science, grounded in SDT, positions students' engagement and science identity as additional components of science learning and persistence. In their framework, the engagement and science identity components act as outcomes, or effects, when the psychological needs of competence, autonomy, and relatedness are met: "When students' intrinsic needs are met in STEM courses, this transforms students' experiences of STEM, their engagement in science, and eventually their own identities as STEM learners" (Skinner et al., 2017, p. 4).

The authors collapsed the three psychological needs highlighted in SDT (competence, autonomy, and relatedness) together to form their first factor, Self-Systems Processes. Thus, this factor is composed of the three short form scales of Competency (the degree to which students feel able to successfully complete coursework), Autonomy/Ownership (how much ownership and personal commitment students feel towards this course), and Relatedness/Belonging (the levels of connection and belonging students feel with their classmates). The remaining two factors in the SPIRES survey, Student Engagement and Identity as a Scientist, were hypothesized to be key contributors to science learning along with the Self-Systems Processes factor. The Student Engagement factor is composed of the four short form scales of Behavioral Engagement, Emotional Engagement, Behavioral Disaffection (the opposite of engagement), and Emotional Disaffection. Finally, the Identity as a Scientist factor is composed of the four short form scales of Science Identity, Science Career Plans, Purpose in Science, and Positive Relationships and Collaborations.

Original Validation of the SPIRES Survey

To validate the SPIRES survey, 1,013 students from eight undergraduate science courses (biology, chemistry, or physics) at an urban university in the Pacific Northwest voluntarily took the survey. The SPIRES survey was administered at two time points online, and students were offered extra credit for their participation. Time 1 was at the end of the first week of an 11 week term and had 856 student participants. Time 2 was at the end of the ninth week of the same 11 week term and had 574 student participants. 417 students participated at both time points.

The authors opted to use confirmatory factor analysis (CFA) instead of exploratory factor analysis (EFA) to validate the 11 scales of the SPIRES survey:

> In contrast to previous measurement work in undergraduate science (e.g. Glynn et al., 2009), which typically uses EFAs to make sense of item pools that were not generated with specific a-priori constructs in mind, CFAs were used in the current study, because all items had been generated or adapted to capture specific theoretically derived constructs, that are consistent with the larger SDT literature. (Skinner et al., 2017, p. 12-13)

First, a CFA of the Time 1 data was performed to confirm and identify the strongest item groupings based on their hypothesized, literature based, scales. Once item groupings were determined, the dimensionality and functioning of these groups were cross-validated with the Time 2 data.

CFA results indicated that nine of the 11 scales in the SPIRES survey were unidimensional at both time points: (a) Emotional Engagement, (b) Science Career Plans, (c) Purpose, (d) Positive Relationships with Peers, (e) Autonomy/Ownership, (f) Behavioral Engagement, (g) Behavioral Disaffection, (h) Emotional Disaffection, and (i) Science Identity. The two scales that were not unidimensional across both time points were Competency and Relatedness/Belonging, which both showed evidence of unidimensionality at Time 1, but not Time 2. Next, the authors tested for reliability using two internal consistency measures, Cronbach's alpha and McDonald's Omega. Behavioral Disaffection was the only scale to demonstrate internal consistency that was not satisfactory across both time points, with Time 1 falling below satisfactory levels but Time 2 falling above satisfactory levels (α = .67 for Time 1 versus α = .76 for Time 2). Finally, a measurement invariance across disciplines (biology, physics, and chemistry) was assessed for all 11 scales at Time 1. These results showed that both Competency and Behavioral Disaffection were the only two scales to show only moderate evidence for invariance across disciplines, while all other scales were either strong or unambiguously invariant across disciplines. To sum, three scales, Competency, Relatedness/Belonging, and Behavioral Disaffection all had "less than ideal measurement properties on at least one of the specific indicators of unidimensionality, reliability, or invariance at one of the time points" (Skinner et al., 2017, p. 18). Despite these results, these three scales were maintained in the final survey.

We are aware of only one independent validity study related to the SPIRES instrument. This study, completed by Komperda et al. (2018), looked at a subset of nine items from the SPIRES survey before the entire survey was published: four items they determined to be related to students' perceived value and five items they determined to be related to belonging. However, this study did not group these items as they are grouped in the final SPIRES survey. Instead, the authors hypothesized that these selected nine items would group into two factors, value and belonging, but CFA did not support their hypothesized factors. The authors argue the value of this type of additional validation effort:

> There is a pervasive misconception that once an instrument has been published in the literature the instrument itself has become validated in some way (Barbera & VandenPlas, 2011) and that data collected from any subsequent administration of the instrument will therefore be equally valid and reliable, even if changes have been made to the wording of the instrument or the context in which it is used. (Komperda et al., 2018, p. 22)

This argument is one that we would like to echo as it encapsulates our reasoning for taking on this independent validity study of the SPIRES survey. We see great value in this survey and its potential to measure students' motivational factors in undergraduate science courses. However, within different university contexts, different student populations, and different time points, psychometric tests have been shown to function differently (e.g., The Science Motivation Questionnaire, (Glynn et al., 2009, 2011). Without our own validity study, inferences made regarding students' science motivational factors from this survey may be unfounded. Thus, the purpose of this study is to perform a cross-validation of the SPIRES survey using a Rasch model, instead of factor analysis, so that institutions, and ourselves, may use this data gathered from this instrument in future research without relying on a sample-dependent factor analysis for each subsequent population this survey is administered.

An Example Validity Study of a Motivational Scale

Validation of motivational scales in science education have largely relied on factor analyses, either exploratory or confirmatory, during early phases of questionnaire development. As an example, consider the exhaustive validation study of the Science Motivation Questionnaire (SMQ; developed by Glynn & Koballa, 2006). The original questionnaire consisted of 30 items that were designed to measure six motivational factors of science students in college: (a) intrinsically motivated science learning, (b) extrinsically motivated science learning, (c) personal relevance of learning science, (d) self-determination (responsibility) for learning science, (e) self-efficacy (confidence) in learning science, and (f) anxiety about science assessment. The original 30 item set was found to be reliable with a Cronbach's alpha of 0.93 (Glynn et al., 2009). However, the exploratory factor analysis (EFA) that was done after the original questionnaire was published to determine construct validity did not support the six factors that the SMQ was hypothesized to measure. Instead, the results of this initial EFA supported five, not six, motivational factors measured by the SMQ: (a) intrinsic motivation and personal relevance, (b) self-efficacy and assessment anxiety, (c) self-determination, (d) career motivation, and (e) grade motivation (Glynn et al., 2009). These EFA results came from a sample of 770 undergraduate non-science majors that were enrolled in a required non-majors science course.

In 2011, the validity of the SMQ was once again studied with another EFA (Glynn et al., 2011). This time, the EFA was conducted from a sample of 367 science majors and 313 non-science majors. The results from this EFA supported five slightly different motivational factors than the EFA published in 2009: (a) intrinsic motivation, (b) self-determination, (c) self-efficacy, (d) career motivation, and (e) grade motivation (Glynn et al., 2011). The final survey, called the SMQ II (Science Motivation Questionnaire II) contains 25 items intended to measure students' motivational factors for learning science (this includes 16 of the original 30 items (Glynn & Koballa, 2006), as well as nine new items (Glynn et al., 2011). However, it is clear from these studies the effect of different samples can have on the construct validity of a new questionnaire. Within different contexts and different students taking the questionnaire at different time points, different factors were supported via EFA.

The validity of the SMQ II has been re-evaluated via both confirmatory and exploratory factor analysis in a variety of contexts (Komperda et al., 2020; Salta & Koulougliotis, 2015; Vasques et al., 2018), which show mixed results regarding construct validity. Taking a different approach, You et al. (2018) used a Rasch analysis to explore the construct and item validity of the SMQ II. Rasch models assume unidimensionality, so this study aimed to validate the SMQ II as an instrument that measures one latent construct of a student's motivational factors. The Rasch analysis demonstrated that the SMQ II was not unidimensional due to two problematic items, which were recommended to be removed from the SMQ II. The authors explained why this result is threatens the validity of the entire measure: “Misfitting items mean a lack of consistency in interpreting the underlying measure, thus, they are not measuring the same latent construct as the rest of the items in the survey” (You et al., 2018, p. 1170). Additionally, 13 of the 25 items on the SMQ II were found to be redundant, as demonstrated by the item-person map, and were recommended to be removed without information loss. Following these results, the authors point out that a Rasch analysis like this is valuable compared to factor analysis because of its ability to assess how well individual items function within the model, and thus the functioning of the instrument as a whole. This Rasch analysis acted as a secondary validation of the SMQ II and was able to parse out important item deficiencies for future iterations of this questionnaire.

Methods

This study utilized a measure validation methodology. DeVellis (2003) describes this process in four stages: (a) planning, (b) construction, (c) quantitative evaluation, and (d) validation. Skinner et al. (2017) conducted the planning and construction phases, and the quantitative evaluation and validations phases began with Komperda et al. (2018) and Skinner et al. (2017). This study expands on the work of Komperda et al. (2018) and further contributes to the developing validity argument for the SPIRES survey by utilizing scores from a new sample of college students, and by using item response theory (IRT). Our research questions were:

1. What are the constructs being measured in Self-determination, Purpose, Identity, and Engagement in Science (SPIRES) scale?
2. Does the data fit the Rasch model and is there adequate reliability for each of the identified dimensions?
3. Is the five-category response scale used appropriately?
4. Is the measure well targeted and which items are the hardest/easiest for the students to endorse?

Validation Approach

In this work, we take an argumentation based approach to instrument validity. Validity is not a static concept in the sense that an instrument can never be “validated” per se. Depending upon the purpose of the instrument and the proposed use of scores resulting from that instrument, a validity argument must be developed (American Educational Research Association et al., 2014; Kane, 2006). Our interest in the construct being measured and in the application of a Rasch model to scores from the instrument are central to this validity argument. In our work, we refer to the three "factors" on the SPIRES as "constructs" which can each have one or more dimensions. This terminology and conceptualization is consistent with a Rasch measurement approach.

Data Collection

Data were collected from introductory biology courses at a mid-sized public urban-serving university in the western United States. 370 students enrolled in undergraduate Biology 1 courses were given the survey on paper during course sessions in Fall 2019. Instructors and researchers did not have access to the data until after the course, and data were de-identified. Data were collected as a part of a larger university Howard Hughes Medical Institute Inclusive Excellence grant (Howard Hughes Medical Institute Award 5200817) addressing equity and attitude changes in undergraduate science courses. Student descriptive statistics are seen in Table 1.

Table 1. Student Demographic Information (n = 370)

Demographic	**Number of Participants (%)**
Age	18-20 = 332 (89.7%) 21-25 = 21 (5.7) 26+ = 17 (4.6%)
Race/Ethnicity	American Indian/Native American/First nations/Indigenous = 9 (2.4%) Asian American or = 82 (22.2%) Black/African American/ African Descent = 24 (6.5%) Hispanic, Latino, or Spanish origin = 130 (35.1%) Pacific Islander = 4 (1.1%) White = 117 (31.6%)
Major	Biology = 176 (47.6%) Other STEM Major = 83 (22.4%) Other non-STEM Major = 111 (30%)
Academic Level	Freshman = 284 (76.8%) Sophomore = 64 (17.3%) Junior = 15 (1.4%) Senior = 2 (0.5%)
Academic Load	Full time student = 298 (80.5%) Part time student = 72 (19.5%)

Analytic Procedure

Examining dimensionality is a vital part of measurement evaluation; we must verify that what is being measured (in the items) actually groups together in the manner it was intended (as the factors). Our data analysis began with principal components analysis of residuals (PCAR) to examine dimensionality (Linacre, 2019). Often referred to as "noise" in the data, residuals are the unexplained part of a correlation. PCAR extracts variance found in the "noise" to examine potential factors. It assumes a single dimension among these residuals and examines if there is enough variance that can explain another possible dimension (Linacre, 1998). A measure is considered unidimensional if no other factor is found in the residuals.

We then built multiple Rasch (Rasch, 1960) models using the Winsteps software (Linacre, 2019). One model was built for each of the 11 factors to examine the structure of the SPIRES survey. To establish unidimensionality, a measure needs to explain 40% or more of the total raw variance, with the first contrast having an eigenvalue of 2.0 or less with less than 5% variance due to the first contrast. We also examined model fit as further evidence of unidimensionality with the mean standardized fit statistics (ZSTD) infit and outfit being close to 0.0 (Linacre, 2019). We then looked at item fit using mean square fit statistics (MNSQ). According to Linacre (2019), MNSQ values between 0.5 – 1.5 are productive of measurement, thus indicating a suitable item. Furthermore, we examined the scales using category probability curves that should indicate an even distribution of the five category options with clearly advancing steps, no evidence of step misfit, and have MNSQ infit values under 2.0 (Linacre, 2019).

Finally, we constructed Wright maps, which present items and persons on the same scale, so that we could examine scale targeting and item difficulty. On these maps, persons are represented as "#" on the left side of the vertical center line, and items are represented by their text abbreviation (e.g., "AUTOE") on the right hand side of the vertical center line. Since the data is coded 5 = strongly agree to 1 = strongly disagree, participants near the top of the left-hand side of these maps are students who have higher levels of endorsement of the construct and students near the bottom have lower endorsement. Items on the right-hand side create a continuum of "easiest" to agree with at the bottom, to "hardest" to agree with items nearer the top.

Validity and Reliability

In item response theory, reliability is measured by calculating person and item separation across the measure. Person separation indicates how well the set of items is able to separate the persons being measured. Item separation indicates how well the sample of persons is able to separate the items used in the survey. A separation value of 2.0 indicates acceptable reliability with higher levels of separation preferred (Linacre, 2019). In our study we also examined construct validity through how the response scales are being used (DeVellis, 2003). We examine this by looking at the ratios between categories, participant scale use, and category structure and function. Additionally, we examined how items group into dimensions in order to assess the internal structure validity.

Results

Dimensionality

Skinner et al. (2017) note the survey is multi-dimensional suggesting three larger constructs with 11 scales.

Therefore, we started by modeling if the data fit as three larger multidimensional constructs: (a) Self-System Processes, (b) Student Engagement, and (c) Identity as a Scientist made up of the 11 individual scales ((a) Competency, (b) Autonomy/Ownership, (c) Relatedness/Belonging, (d) Behavioral Engagement, (e) Emotional Engagement, (f) Behavioral Disaffection, (g) Emotional Disaffection, (h) Science Identity, (i) Science Career Plans, (j) Purpose in Science, and (k) Positive Relationships and Collaborations). Recall that we are referring to the three larger "factors" on the SPIRES as "constructs", and each of those constructs can have one or more dimensions.

Self-System Process was considered a multi-dimensional construct (1st contrast eigenvalue = 2.84). Items comprising this construct were all the items from the Competency, Autonomy/Ownership, and Relatedness/Belonging scales. Therefore, each of these scales are considered to be *dimensions* of the Self-Systems Processes construct.

Student Engagement was also considered a multi-dimensional construct (1st contrast eigenvalue = 4.01) with items clearly grouping into Behavior and Emotional Engagement or Disaffection separately. When modeling the behavior items only these items clearly grouped into either Behavioral Engagement or Behavioral Disaffection. The same pattern was found for Emotional items, with the items grouping separately into Emotional Engagement or Emotional Disaffection. Therefore, these four scales are considered to be four *dimensions* of the Student Engagement construct.

Identity as a Scientist was also multi-dimensional (1st contrast eigenvalue = 3.57). The same pattern was found with Science Identity, Science Career Plans, and Purpose in Science, with items grouping into those three distinct dimensions. The Positive Relationships and Collaborations scale was also found to be unidimensional (1st contrast eigenvalue = 1.65). *The results of our dimensionality tests mean that we will discuss the 11 dimensions found above separately, and we will group the final models by the four larger constructs (or "big ideas").*

Self-System Processes Models

For the Self-System Process construct, the data fit the Rasch models for the three dimensions of Autonomy, Competency, and Relatedness (see Table 2). The Relatedness dimension showed a possible second dimension with the eigenvalue of the first contrast right at 2.00 and explaining 18.6% of the variance while the Rasch dimension's items only explain 15.1% of the total variance. Items in this possible sub-dimension include two reverse coded items, but removing them did not improve the model and left this dimension with only three items. All items fit the model well with mean square (MNSQ) values in the recommended range and all items correlate to the total measure above $r = 0.61$ (see Table 3).

Reliability is strong with Cronbach's alphas ranging from 0.77 – 0.82. Reliability of person separation is interpreted in a similar manner to Cronbach's alpha with values above 0.70 considered reliable (Linacre, 2019). These values indicate how spread out, and thus reliable, the persons and items are for the scale. All three dimensions demonstrate reliability with person separation values between 1.84 - 2.17, reliability of persons between 0.79 - 0.83, and reliability of items between 0.96 - 1.00 (see Table 2).

Table 2. Self-System Process Dimensionality, Fit, and Separation

Index	**Autonomy**	**Competency**	**Relatedness**
Dimensionality – eigenvalue for 1st contrast	1.83	1.78	2.00
Mean ZSTD Infit	0.00	-0.10	-0.20
SD ZSTD Infit	1.20	1.30	1.30
Mean ZSTD Outfit	-0.10	-0.10	-0.20
SD ZSTD Outfit	1.10	1.30	1.30
Model Person Separation	1.84	2.17	1.97
Model Person Root Mean Square Error	0.78	0.76	0.70
Model Reliability of Person Separation	0.77	0.82	0.79
Cronbach's Alpha	0.79	0.82	.083
Model Item Separation	17.75	4.66	5.01
Model Reliability of Item Separation	1.00	0.96	0.96

Note. ZSTD Infit is a *t* statistic testing model fit with sensitivity to midrange observations. ZSTD Outfit is a *t* statistic testing model fit with sensitivity to extreme responses. Person/Item Separation is the ratio of the true standard deviation to the error standard deviation. Person Root Mean Square Error is standard error of the measure inflated for misfit. Reliability of Person/Item Separation = Separation² / (1 + Separation²).

Table 3. Self-System Process Item Fit Statistics

Item#	Logit Position	*SE*	Infit MNSQ	Pt-Measure Correlation
Autonomy				
A - I want to understand the subject	-1.43	0.10	0.73	0.69
B - I want to learn new things	-0.79	0.09	0.58	0.78
C - Doing well in science is important to me	-1.54	0.10	0.83	0.67
D - It's fun to answer challenging science questions	1.24	0.08	1.10	0.75
E - I just learn the stuff I have to in order to pass the test(s)	2.52	0.08	1.45	0.69
Competency				
A - I am good at science	0.19	0.08	0.81	0.80
B - I find it easy to understand the things we are learning in this class	0.58	0.08	0.90	0.80
C - Even if they are challenging, I can do well in my science classes	0.07	0.08	0.76	0.80
D - I don't have the intelligence/brains to succeed in science	-0.64	0.09	1.05	0.71
E - When I do poorly in a science course, I usually can't figure out why	-0.20	0.09	1.47	0.61
Relatedness				
A - This course is a good place for students like me	-0.29	0.08	0.83	0.76

Item#	Logit Position	*SE*	Infit MNSQ	Pt-Measure Correlation
B - This is the right course for me to be taking now	-0.65	0.09	0.86	0.76
C - In science courses, I feel like an outsider	0.25	0.08	0.83	0.77
D - I fit in well with the other students in this class	0.45	0.08	1.11	0.70
E - I'm not really sure that science is the right major for me	0.24	0.08	1.36	0.73

Note. Item logit position is the value seen in the Wright Map which creates the construct's continuum. SE is standard error of the logit position. MNSQ is a statistic testing model fit that removes extreme observations. Pt-Measure correlation is the relationship between the individual item and the total measure.

The 5-point scale of strongly agree to strongly disagree was mostly used as intended. All categories advanced in order with even step structures and no misfit (see Table 4). The only concern is that categories 2 (disagree) and 1 (strongly disagree) are somewhat underrepresented in the responses, as they were only selected by 8-10% and 2-3% respectively by students (see Figure 1).

Table 4. Self-System Process Scale Step Structure

Category	Observed Percentage	Observed Average	Infit MNSQ	Step Structure
Autonomy				
1 – *strongly disagree*	3	-1.74	2.00	(-4.15)
2 - dis*agree*	10	-1.28	0.90	-1.98
3 – *neutral*	17	0.41	0.93	-0.14
4 - *agree*	35	2.07	1.02	1.96
5 – *strongly agree*	34	4.18	0.94	(4.43)
Competency				
1 – *strongly disagree*	2	-1.66	0.93	(-4.04)
2 - dis*agree*	10	-0.69	1.05	-2.08
3 – *neutral*	24	0.12	0.91	-0.39
4 - *agree*	44	1.70	0.96	2.03
5 – *strongly agree*	20	6.64	1.10	(4.86)
Relatedness				
1 – *strongly disagree*	3	-1.32	1.18	(-3.38)
2 - dis*agree*	8	-0.69	1.04	-1.75
3 – *neutral*	24	0.25	0.87	-0.33
4 - *agree*	39	1.45	0.96	1.69
5 – *strongly agree*	25	2.90	1.06	(4.09)

Note. Observed percentage is the percent of all responses for that category. Observed average is the average of the measure to produce the responses observed in the category. Infit MNSQ is the average of the infit MNSQs associated with responses in that category. Step Structure is the logit position where the conditional probability of being in either category is equal.

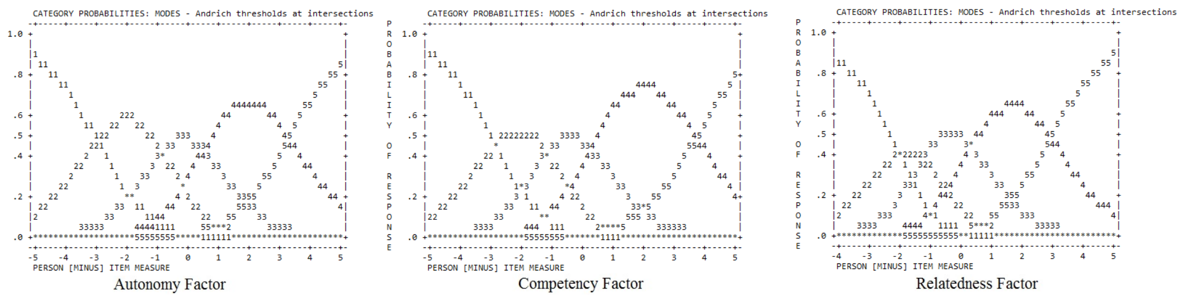


Figure 1. Self-Systems Process Category Probability Curves

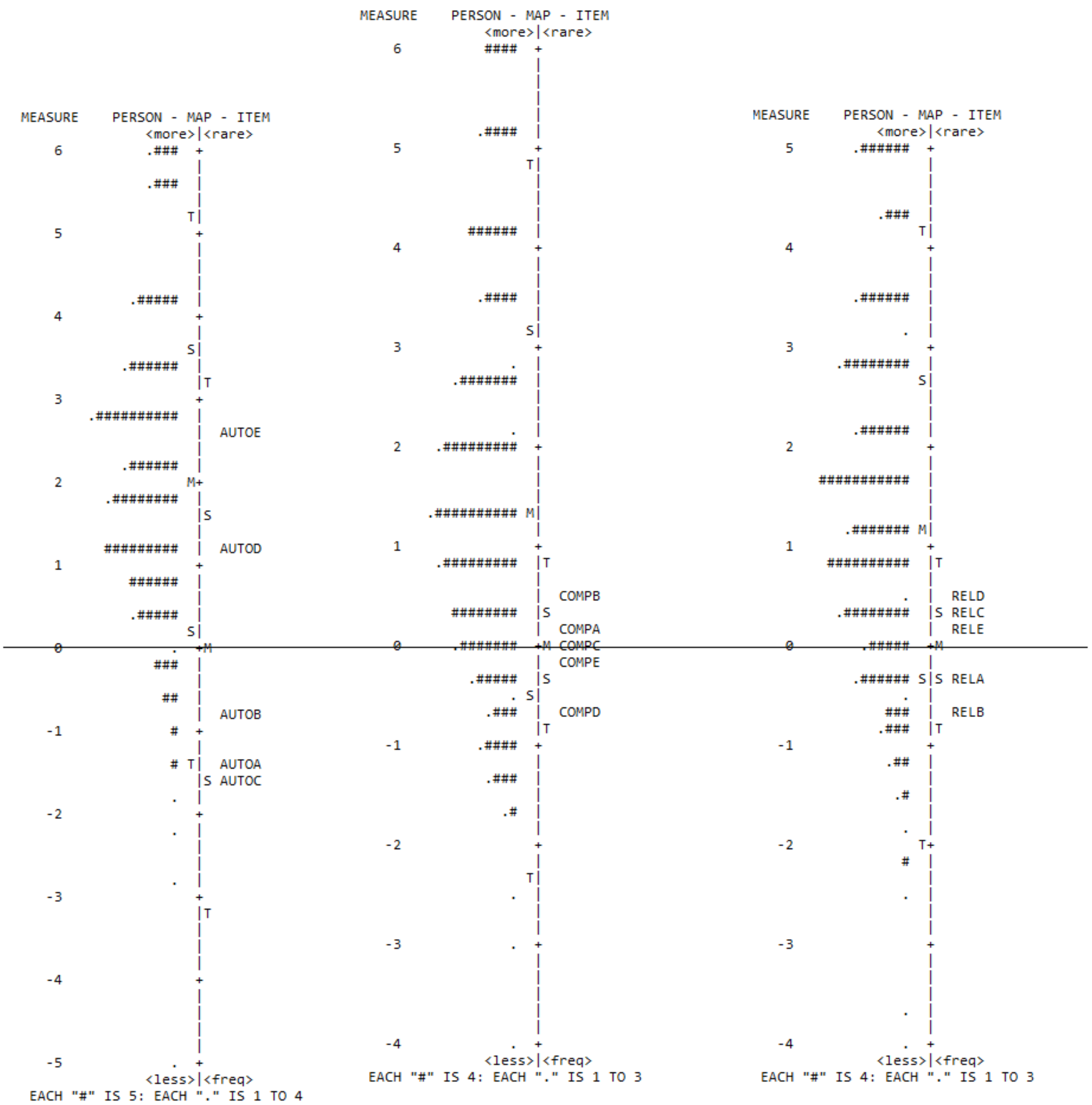


Figure 2. Wright Maps for Self-System Process

[Note: Persons are represented as "#" on the left side of the vertical center line, and items are represented by their text abbreviation (e.g., "AUTOA") on the right hand side of the vertical center line. Since the data is coded 5 = strongly agree to 1 = strongly disagree, participants near the top of the left-hand side of these maps are students who have higher levels of endorsement of the construct and students near the bottom have lower endorsement. Items on the right-hand side create a continuum of "easiest" to agree with at the bottom, to "hardest" to agree with items nearer the top.]

Wright maps for these three dimensions are seen in Figure 2 with logit positions in Table 3. The Competency and Relatedness/Belonging dimension have items groups in the center only covering logit positions between +/- 1. While this demonstrates good spread, it is not well targeted for roughly half the sample, which means that half of students are so agreeable to these dimensions, they are not captured well by the item ruler. Autonomy/Ownership is spread between +/- 2 logit position and has two items that extend into the students with higher endorsement.

Student Engagement Models

For the Student Engagement construct, the data fit the models well for the dimensions of Behavioral Engagement, Behavioral Disaffection, Emotional Engagement, and Emotional Disaffection (see Table 5). Person separation was lower than desired for both behavioral dimensions at 1.53 for Behavioral Engagement and 1.56 for Behavioral Disaffection. This is seen in the Wright Maps as well with students grouped closely together in the positive endorsement range and not spread ideally throughout the continuum. Still, reliability of persons was strong ranging 0.70 – 0.83, as were the Cronbach's alphas between 0.76 – 0.86. Item reliability was strong ranging between 0.92 – 0.99. All items fit well and item-total correlations ranged from 0.58 – 0.81 (see Table 6).

Table 5. Engagement and Disaffection Dimensionality, Fit, and Separation

Index	Behavioral Engagement	Behavioral Disaffection	Emotional Engagement	Emotional Disaffection
Dimensionality – eigenvalue for 1st contrast	1.53	1.80	1.59	1.67
Mean ZSTD Infit	-.020	-0.20	-0.10	-0.10
SD ZSTD Infit	1.40	1.30	1.20	1.40
Mean ZSTD Outfit	-0.20	-0.10	-0.10	-0.20
SD ZSTD Outfit	1.40	1.30	0.98	1.40
Model Person Separation	1.53	1.56	2.16	2.24
Model Person Root Mean Square Error	0.87	0.75	0.89	0.75
Model Reliability of Person Separation	0.70	0.71	0.82	0.83
Cronbach's Alpha	0.82	0.76	0.86	0.85
Model Item Separation	3.50	5.04	8.90	10.88
Model Reliability of Item Separation	0.92	0.96	0.99	0.99

Table 6. Engagement and Disaffection Item Fit Statistics

Item#	Logit Position	*SE*	Infit ZSTD	Pt-Measure Correlation
Behavioral Engagement				
A - I pay attention in class	0.01	0.11	1.06	0.71
B - I study for this class	0.55	0.10	0.89	0.76
C - I try hard to understand the professor's lectures	-0.36	0.11	1.01	0.72
D - I keep up with work for this class	0.31	0.10	1.10	0.75
E - I try hard to do well in this class	-0.51	0.12	0.94	0.76

Item#	Logit Position	*SE*	Infit ZSTD	Pt-Measure Correlation
Behavioral Disaffection				
A - It's hard to make myself come to this class	0.11	0.08	1.37	0.61
B - Outside of class, I don't put much work in on this course	-.16	0.08	0.62	0.77
C - Anything I do for this class is always last minute	-.54	0.08	0.90	.075
D - I don't really study for this class	0.00	0.09	0.74	0.74
E - I work on other things when I'm in this class	-0.81	0.10	1.36	0.58
Emotional Engagement				
A - I enjoy the time I spend in this class	-0.11	0.11	0.99	0.80
B - The material we cover is interesting	-0.64	0.11	0.90	0.81
C - It's exciting to make connections between the ideas learned in this class	-0.68	0.11	0.84	0.81
D - The material we cover in class is challenging (in a good way)	-0.43	0.11	1.05	0.77
E - The readings for this class are interesting	1.87	0.10	1.21	0.77
Emotional Disaffection				
A - When in class, I feel bored	-0.13	0.09	0.92	0.76
B - This class is stressing me out	1.75	0.08	1.60	0.74
C - This class can be pretty dull	-0.20	0.09	0.80	0.80
D - When I'm in this class, I can't wait for it to be over	-0.31	0.09	0.66	0.81
E - This class is no fun	-1.11	0.09	0.80	0.76

While all scale fit statistics were acceptable with no evidence of misstep (see Table 7), scale category probability curves for the behavioral dimensions were concerning. Figure 3 shows strongly muted category 3 (neutral) and category 2 (disagree). It is likely due to low responses in category 2 and 1 across all four dimensions.

Table 7. Engagement and Disaffection Scale Step Structure

Category	Observed Percentage	Observed Average	Infit MNSQ	Step Structure
Behavioral Engagement				
1 – *strongly disagree*	0	0.23	1.44	(-3.45)
2 - di*sagree*	2	0.05	1.00	-1.90
3 – *neutral*	10	0.84	1.06	-0.61
4 - *agree*	45	1.94	0.88	1.79
5 – *strongly agree*	43	3.93	0.98	(4.76)
Behavioral Disaffection				
1 – *strongly disagree*	1	-0.26	1.20	(-3.60)
2 - di*sagree*	6	-0.01	1.06	-1.65
3 – *neutral*	14	0.53	0.93	-0.27
4 - *agree*	46	1.59	0.88	1.58
5 – *strongly agree*	34	3.13	1.05	(4.21)

Category	Observed Percentage	Observed Average	Infit MNSQ	Step Structure
Emotional Engagement				
1 – *strongly disagree*	1	-3.32	1.48	(-5.82)
2 - dis*agree*	4	-1.05	1.29	-3.28
3 – *neutral*	24	0.86	1.03	-0.26
4 - *agree*	45	3.00	0.92	3.28
5 – *strongly agree*	25	5.47	0.90	(6.31)
Emotional Disaffection				
1 – *strongly disagree*	5	-2.47	1.26	(-3.97)
2 - dis*agree*	11	-1.35	0.91	-2.08
3 – *neutral*	22	0.08	0.93	-0.43
4 - *agree*	42	1.72	0.91	2.03
5 – *strongly agree*	20	3.57	1.08	(4.85)

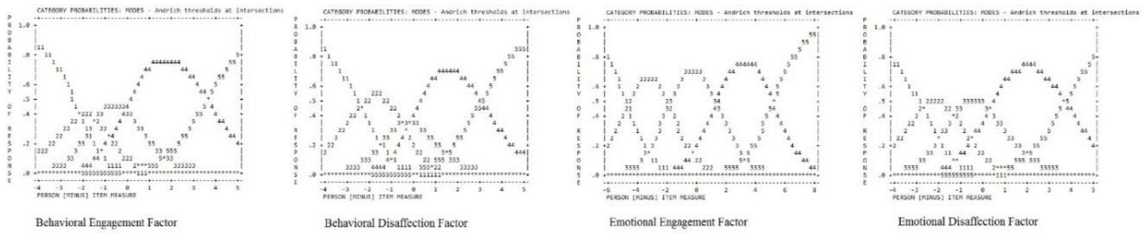


Figure 3. Engagement and Disaffection Category Probability Curves

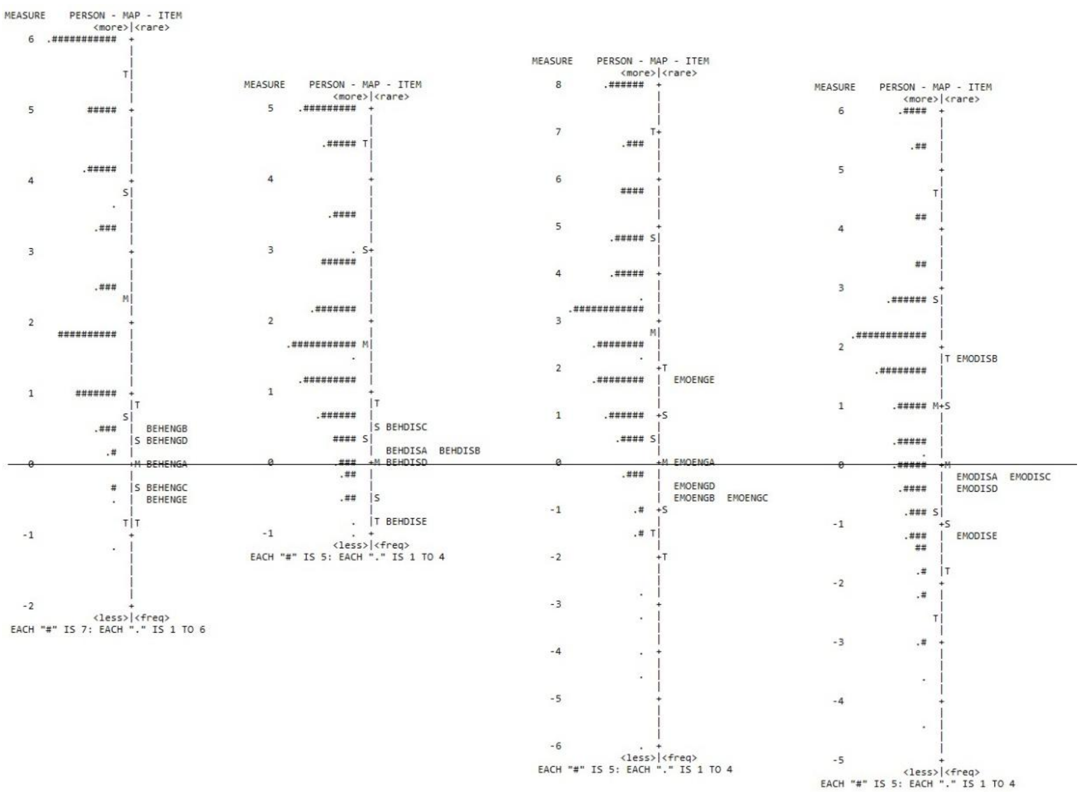


Figure 4. Wright Map for Engagement and Disaffection

Wright maps for the four dimensions show these are poorly targeted to the students as the item ruler does not measure where students are grouped (see Figure 4; Table 6). While both emotional dimensions have items that reach where more students' endorsement is, the behavioral items are grouped too closely together and not ideal for this sample of students. As the disaffection items are reverse coded, these maps show students grouped towards the top indicating high engagement and low disaffection endorsement.

Identity as a Scientist Models

For the Identity as a Scientist construct, the data fit the Rasch model well (Table 8). Science Career Plans showed a second, sub-dimension (eigenvalue = 2.07) with the variance explaining more for this contrast (15.4%) than the variance of items in the first contrast (8.4%). The reverse coded item D ("I'm just not cut out for a career in science") was the only item in the first contrast and showed misfit with MNSQ = 1.99 (Table 9). The same was seen in the Purpose in Science dimension where the first contrast also explains more (18.8%) than the items in the first dimension (5.4%). Item D ("Sometimes I feel like I don't belong in science"), also reverse coded, showed misfit with MNSQ = 1.53. Removing these items did not improve dimensionality or fit and left only 3 items in each scale, which is not ideal for measurement. All statistics reported include all four Science Career Plans and Purpose in Science items.

Table 8. Identity as a Scientist Dimensionality, Fit, and Separation

Index	Career Plans	Purpose in Science	Science Identity
Dimensionality – eigenvalue for 1st contrast	2.07	1.75	1.98
Mean ZSTD Infit	-0.20	-0.20	-0.20
SD ZSTD Infit	1.30	1.20	1.30
Mean ZSTD Outfit	-0.20	-0.20	-0.20
SD ZSTD Outfit	1.30	1.20	1.30
Model Person Separation	2.31	1.70	2.21
Model Person Root Mean Square Error	0.97	1.22	0.80
Model Reliability of Person Separation	0.84	0.74	0.83
Cronbach's Alpha	0.92	0.92	0.87
Model Item Separation	5.74	4.27	10.53
Model Reliability of Item Separation	0.97	0.95	0.99

Table 9. Identity as a Scientist Item Fit Statistics

Item#	Logit Position	*SE*	Infit ZSTD	Pt-Measure Correlation
Career Plans				
A - For the career I want, I need a degree in science	0.04	0.11	1.13	0.87
B - I am planning on a job that involves science	-0.45	0.12	0.58	0.91
C - Science is important for my future career	-0.64	0.12	0.56	0.90
D - I'm just not cut out for a career in science	1.06	0.10	1.70	0.79

Item#	Logit Position	*SE*	Infit ZSTD	Pt-Measure Correlation
Purpose in Science				
A - Science can help solve many of society's problems	0.77	0.14	0.99	0.87
B - I believe that science can help make the world a better place	-0.42	0.16	0.77	0.89
C - I can see lots of ways that science makes a positive difference in our everyday live	-0.86	0.16	0.61	0.90
D - If everyone in our society learned more about science, we could all make better decisions about important things like politics, medicine, and the environment	0.52	.015	1.52	0.83
Science Identity				
A - I am the kind of person who can succeed in science	-0.53	0.10	0.82	0.78
B - I think science is fascinating	-1.52	0.11	1.23	0.67
C - I feel at home in science	0.64	0.09	0.80	0.84
D - Sometimes I feel like I don't belong in science	1.44	0.09	1.15	0.83
E - I don't think I could ever really feel comfortable in science	-0.03	0.09	0.93	0.79

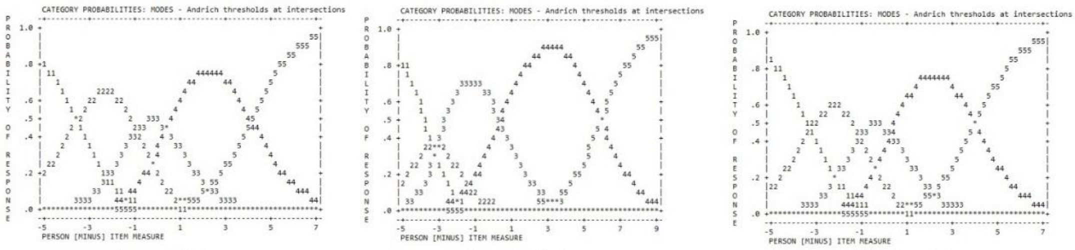
Purpose in Science showed less person separation than desired at 1.70, but Science Career Plans and Science Identity had acceptable separation at 2.21-2.31. Still, reliability for person separation was strong ranging from 0.74 – 0.84. Cronbach's alphas were also strong ranging from 0.87 – 0.90, as was reliability of item separation ranging from 0.94 – 0.99 (Table 8).

While no misfit or misstep was found, scale use was concerning in regards to the neutral, disagree, and strongly disagree categories (Table 10; Figure 5). Neutral was only chosen 10% of the time for Career Plans and Purpose in Science and 18% for Science Identity. Disagree was chosen 6% for Career Plans, 1% for Purpose in Science, and 10% for Science Identity. Strongly disagree was chosen 3% for Science Career Plans, 0% for Purpose in Science, and 2% for Science Identity.

Table 10. Identity as a Scientist Scale Step Structure

Category	Observed Percentage	Observed Average	Infit MNSQ	Step Structure
Career Plans				
1 – *strongly disagree*	3	-3.12	2.22	(-4.73)
2 – di*sagree*	6	-1.46	1.08	-2.32
3 – *neutral*	10	0.56	1.08	-0.25
4 – *agree*	33	2.39	0.95	2.29
5 – *strongly agree*	48	4.91	0.82	(5.24)
Purpose in Science				
1 – *strongly disagree*	0	0.05	2.34	(-4.79)
2 – di*sagree*	1	0.92	1.34	-3.21
3 – *neutral*	10	0.30	0.96	-1.21

Category	Observed Percentage	Observed Average	Infit MNSQ	Step Structure
4 – *agree*	39	3.13	0.86	3.19
5 – *strongly agree*	49	6.43	0.90	(7.09)
Science Identity				
1 – *strongly disagree*	2	-2.46	1.00	(-4.41)
2 – dis*agree*	10	-1.02	1.02	-2.23
3 – *neutral*	18	0.14	0.85	-0.36
4 – *agree*	41	2.10	0.81	2.19
5 – *strongly agree*	29	4.13	1.30	(5.14)



Autonomy Factor Competency Factor Relatedness Factor

Figure 5. Identity as a Scientist Category Probability Curves

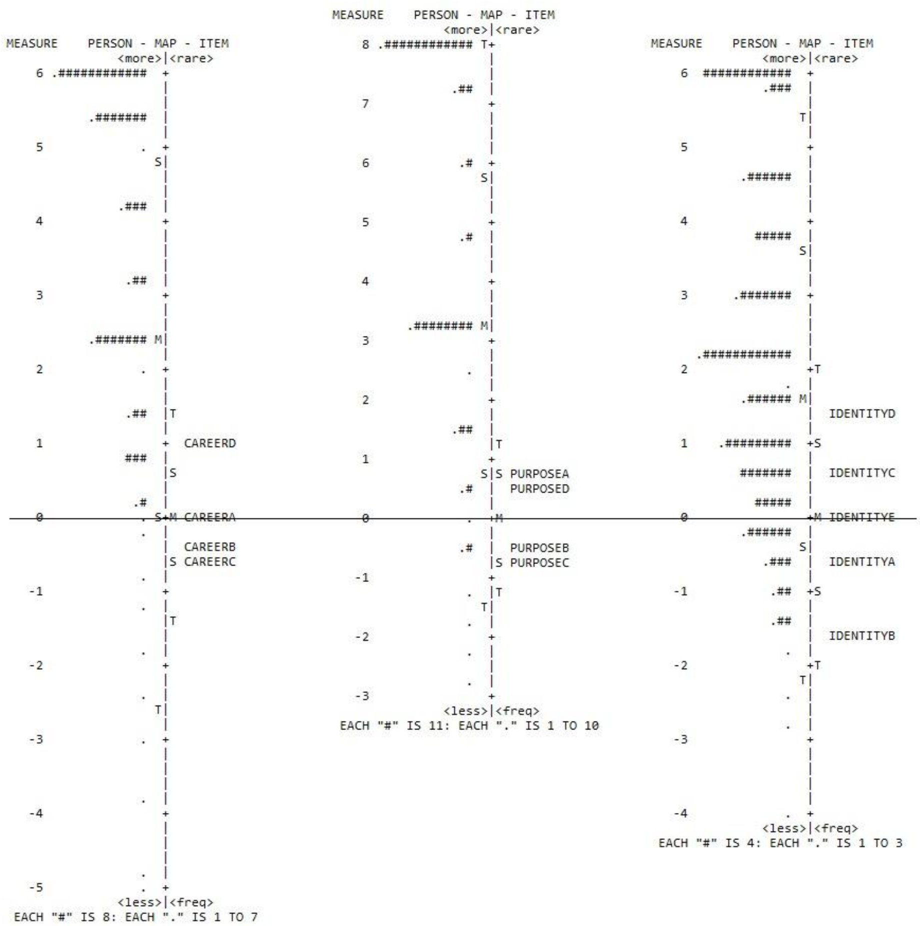


Figure 6. Wright Map for Identity as a Scientist

Wright maps show the same heavy endorsement for the three dimensions (Figure 6). Science Career and Purpose in Science items are spread between +/-1 logit positions but are not measuring this student population well. Science Identity items are spread nicely to capture those with less endorsement of identity, but not those with higher endorsement.

Positive Relationships and Collaborations Model

The Positive Relationships and Collaborations construct is the only larger concept in the SPIRES measure that is a single, unidimensional big idea. The data fit the Rasch model well with no dimensionality or item fit concerns (see Table 11; Table 12). There is good separation and strong reliability for persons and items.

Table 11. Relationship with Peers Dimensionality, Fit, and Separation

Index	Relationship with Peers
Dimensionality – eigenvalue for 1st contrast	1.65
Mean ZSTD Infit	-0.20
SD ZSTD Infit	1.30
Mean ZSTD Outfit	-0.20
SD ZSTD Outfit	1.30
Model Person Separation	2.20
Model Person Root Mean Square Error	0.87
Model Reliability of Person Separation	0.83
Cronbach's Alpha	0.91
Model Item Separation	6.66
Model Reliability of Item Separation	0.98

Table 12. Relationship with Peers Item Fit Statistics

Item#	Logit Position	*SE*	Infit ZSTD	Pt-Measure Correlation
A - I have gotten to know other students in this class	-0.48	0.10	0.87	0.88
B - In this class, I have found people to study with	0.35	0.10	0.80	0.90
C - In this class, I know people I could ask for help with assignments	-0.78	0.10	0.86	0.88
D - Some students from this class and I are thinking about taking another course together	0.91	0.10	1.32	0.86

Scale statistics are acceptable, but neutral is muted in the category probability curves and the two disagree categories are not highly chosen (see Table 13; Figure 7). The Wright map shows items are spread to measure those with mid-lower levels of endorsement well but not those with high endorsement of the construct (see Figure 8).

Table 13. Relationship with Peers Scale Step Structure

Category	Observed Percentage	Observed Average	Infit MNSQ	Step Structure
1 – *strongly disagree*	5	-2.84	2.42	(-5.24)
2 - dis*agree*	13	-1.34	0.83	-2.34
3 – *neutral*	15	0.50	0.86	0.02
4 - *agree*	34	2.11	0.81	2.34
5 – *strongly agree*	32	3.86	1.10	(5.20)

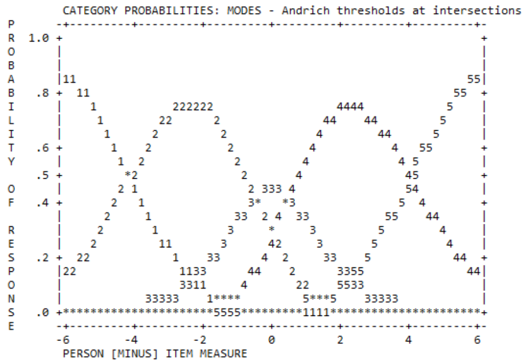


Figure 7. Relationship with Peers Category Probability Curves

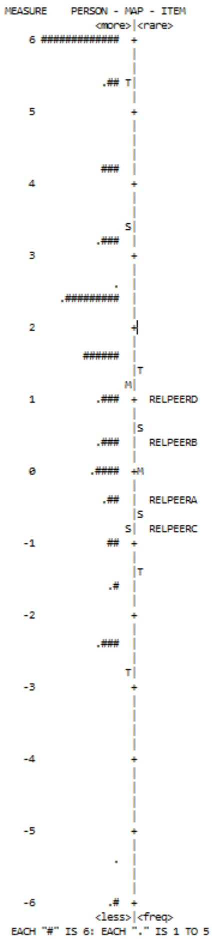


Figure 8. Wright Map for Relationship with Peers

Discussion

In this study, we used the Rasch model to cross-validate the Self-determination, Purpose, Identity, and Engagement in Science (SPIRES) survey (Skinner et al., 2017), a psychometric instrument developed to measure science students' motivational traits. The SPIRES survey was originally validated via confirmatory factor analysis by (Skinner et al., 2017), which showed three psychometric factors:

(a) Self-System Processes,

(b) Student Engagement, and

(c) Identity as a Scientist.

Considering that factor analysis is sample-dependent, these results are reasonable within the student population from which the SPIRES was originally tested and validated. However, to use this survey within a new student population or educational context, instrument functioning, validity, and reliability needs to be demonstrated independently (Kane, 2006; Komperda et al., 2018). Continually relying on factor analysis to validate a psychometric survey like the SPIRES can produce inconsistent results (i.e., Glynn et al., 2009, 2011) as these instruments often function differently within distinct university contexts, student populations, and time points. Therefore, we took a Rasch modeling approach, instead of factor analysis which is sample-independent, to cross-validate this survey. In doing so, we provide evidence for a validity argument that other researchers can use in future work without relying on a sample-dependent factor analysis for each subsequent population this survey is administered.

Our Rasch modeling results indicate that the SPIRES survey is psychometrically robust and is composed of four constructs, three of which are multi-dimensional (Self-Systems Processes, Student Engagement, and Identity as a Scientist), and one of which is unidimensional (Positive Relationships and Collaborations). The Self-Systems Processes, Student Engagement, and Identity as a Scientist constructs are collectively comprised of 10 dimensions, designated by the 10 scales of (a) Competency, (b) Autonomy/Ownership, (c) Relatedness/Belonging, (d) Behavioral Engagement, (e) Emotional Engagement, (f) Behavioral Disaffection, (g) Emotional Disaffection, (h) Science Identity, (i) Science Career Plans, and (j) Purpose in Science. The fourth construct, which is unidimensional, is composed of the 11th scale: Positive Relationships and Collaborations. This result differs from the original factor analysis validation of the SPIRES survey done by Skinner et al. (2017), which suggested that the SPIRES survey consisted of only three factors (or constructs, in our validation language): (a) Self-System Processes, (b) Student Engagement, and (c) Identity as a Scientist. Our analysis indicated that these three constructs are mutli-dimensional and should not be treated as the sole constructs of this survey. Rather, the 11 scales of the SPIRES are clear individual dimensions of these constructs and should be treated as such.

Our work also demonstrates the advantage of using a Rasch modeling approach over factor analysis for instrument validation. Different factors or a different number of factors could be found *every time* the instrument is validated using factor analysis in distinct student populations, such as in (Glynn et al., 2009, 2011). The original validation of the SPIRES suggested the survey is composed of three larger ideas or constructs, while our Rasch modeling approach suggests there are four constructs. However, since our analyses were sample-independent, we conclude that the SPIRES survey is a four construct survey and may be treated as such across educational contexts without further need to validate using factor analysis, which could continuously produce inconsistent results.

Conclusion

Researchers interested in using the SPIRES to examine science students' motivational traits with a Self Determination Theory framework can use our work as a basis for their validity argument for the instrument. However, we strongly advise researchers to consider three of the SPIRES constructs (Self-Systems Processes, Student Engagement, Identity as a Scientist) as multi-dimensional, and one construct (Positive Relationships and Collaborations) as unidimensional when analyzing their data. This has implications for use of each of the 11 scales, or use of the holistic measure. For example, when considering the relationship between these dimensions as they might relate to students' intersectional identities, any unwarranted assumptions about dimensionality would mask the potential complexity of these constructs. As researchers continue to work on understanding intersectionalities, this type of complexity needs to be elucidated as much as possible.

In summary, we recommend any future analysis of data from the SPIRES survey to consider each of the 11 scales as dimensions of these four constructs. However, we also recommend using this survey as originally published without editing or removing any of the items within the 11 scales, as removing any slightly misfitting items in any of the scales did not improve model fit for each construct in our analysis. Future psychometric research could examine scale use more and how removing the "neutral" option may impact scale functioning, as this option was muted in several of the scales. Additionally, a very important future line of research would include a Differential Item Functioning (DIF) analysis to examine how specific items within the SPIRES survey may function differently across demographic and social identifier groups.

References

American Educational Research Association, American Psychological Association, National Council on Measurement in Education, & Joint Committee on Standards for Educational and Psychological Testing (U.S.). (2014). *Standards for Educational and Psychological Testing*. American Educational Research Association. https://market.android.com/details?id=book-clI_mAEACAAJ

Barbera, J., & VandenPlas, J. R. (2011). All Assessment Materials Are Not Created Equal: The Myths about Instrument Development, Validity, and Reliability. In *Investigating Classroom Myths through Research on Teaching <break></break>and Learning* (Vol. 1074, pp. 177–193). American Chemical Society. https://doi.org/10.1021/bk-2011-1074.ch011

Carlone, H. B., & Johnson, A. (2007). Understanding the science experiences of successful women of color: Science identity as an analytic lens. *Journal of Research in Science Teaching*, *44*(8), 1187–1218. https://doi.org/10.1002/tea.20237

Deci, E. L., & Ryan, R. M. (2000). The "What" and "Why" of Goal Pursuits: Human Needs and the Self-Determination of Behavior. *Psychological Inquiry*, *11*(4), 227–268. https://doi.org/10.1207/S15327965PLI1104_01

DeVellis, R. F. (2003). *Scale development: Theory and applications*. SAGE Publications.

Glynn, S. M., & Koballa, T. R., Jr. (2006). Motivation to learn in college science. In J. J. Mintzes & W. H. Leonard (Eds.), *Handbook of college science teaching* (pp. 25–32). National Science Teachers Association Press.

Glynn, S. M., Brickman, P., Armstrong, N., & Taasoobshirazi, G. (2011). Science motivation questionnaire II: Validation with science majors and nonscience majors. *Journal of Research in Science Teaching*, *48*(10), 1159–1176. https://doi.org/10.1002/tea.20442

Glynn, S. M., Taasoobshirazi, G., & Brickman, P. (2009). Science Motivation Questionnaire: Construct validation with nonscience majors. *Journal of Research in Science Teaching*, *46*(2), 127–146. https://doi.org/10.1002/tea.20267

Huvard, H., Talbot, R. M., Mason, H., Thompson, A. N., Ferrara, M., & Wee, B. (2020). Science identity and metacognitive development in undergraduate mentor-teachers. *International Journal of STEM Education, 7*, 1-17.

Jack, B. M., Lin, H.-S., & Yore, L. D. (2014). The synergistic effect of affective factors on student learning outcomes. *Journal of Research in Science Teaching*, *51*(8), 1084–1101. https://doi.org/10.1002/tea.21153

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational Measurement* (pp. 17–64, 4th ed). American Council on Education and Praeger. http://www.mendeley.com//research/validation-12//

Knekta, E., Chatzikyriakidou, K., & McCartney, M. (2020). Evaluation of a Questionnaire Measuring University Students' Sense of Belonging to and Involvement in a Biology Department. *CBE—Life Sciences Education*, *19*(3), ar27. https://doi.org/10.1187/cbe.19-09-0166

Komperda, R., Hosbein, K. N., & Barbera, J. (2018). Evaluation of the influence of wording changes and course type on motivation instrument functioning in chemistry. *Chemistry Education Research and Practice*, *19*(1), 184–198. https://doi.org/10.1039/C7RP00181A

Komperda, R., Hosbein, K. N., Phillips, M. M., & Barbera, J. (2020). Investigation of evidence for the internal structure of a modified science motivation questionnaire II (mSMQ II): a failed attempt to improve instrument functioning across course, subject, and wording variants. In *Chemistry Education Research and Practice* (Vol. 21, Issue 3, pp. 893–907). https://doi.org/10.1039/d0rp00029a

Linacre, J. M. (1998). Structure in Rasch residuals: why principal components analysis. *Rasch Measurement Transactions*, *12*(2), 636.

Linacre, J. M. (2019). *A User's Guide to Winsteps Ministep: Rasch Model Computer Programs* (Version 4.4.5). winsteps.com

Lomax, R. G., & Schumacker, R. E. (2010). *A beginner's guide to structural equation modeling*. Routledge.

National Science Foundation, National Center for Science and Engineering Statistics. (2019). *Women, minorities, and persons with disabilities in science and engineering: 2019* (Special Report NSF 19-304). Alexandria, VA.

PCAST (President's Council on Science and Technology). (2012). *Engage to Excel*. Washington, DC: President's Council of Advisors on Science and Technology.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Danish Institute for Educational Research.

Riegle-Crumb, C., King, B., & Irizarry, Y. (2019). Does STEM stand out? Examining racial/ethnic gaps in persistence across postsecondary fields. *Educational Researcher*, *48*(3), 133–144.

Romine, W. L., Walter, E. M., Bosse, E., & Todd, A. N. (2017). Understanding patterns of evolution acceptance-A new implementation of the Measure of Acceptance of the Theory of Evolution (MATE) with Midwestern university students. *Journal of Research in Science Teaching*, *54*(5), 642–671. https://doi.org/10.1002/tea.21380

Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness* (Vol. 756). Guilford Press. https://psycnet.apa.org/fulltext/2017-04680-000.pdf

Salta, K., & Koulougliotis, D. (2015). Assessing motivation to learn chemistry: adaptation and validation of Science Motivation Questionnaire II with Greek secondary school students. *Chemistry Education Research and Practice*, *16*(2), 237–250. https://doi.org/10.1039/C4RP00196F

Skinner, E., Saxton, E., Currie, C., & Shusterman, G. (2017). A motivational account of the undergraduate experience in science: Brief measures of students' self-system appraisals, engagement in coursework, and identity as a scientist. *International Journal of Science Education*, *39*(17), 2433–2459. https://doi.org/10.1080/09500693.2017.1387946

Stets, J. E., Brenner, P. S., Burke, P. J., & Serpe, R. T. (2017). The science identity and entering a science occupation. *Social Science Research*, *64*, 1–14. https://doi.org/10.1016/j.ssresearch.2016.10.016

Tsai, C.-Y., Li, Y.-Y., & Cheng, Y.-Y. (2017). The Relationships Among Adult Affective Factors, Engagement in Science, and Scientific Competencies. *Adult Education Quarterly*, *67*(1), 30–47. https://doi.org/10.1177/0741713616673148

Vasques, D. T., Yoshida, L., & Ellinger, J. (2018). Validity and reliability of the science motivation questionnaire II (SMQ II) in the context of a japanese university. *Conference*. https://books.google.com/books?hl=en&lr=&id=QJpaDwAAQBAJ&oi=fnd&pg=PA80&dq=SMQ+II+validation&ots=69oecqoDyE&sig=33qjIYX38vKM-K4v3jcdrmMzvHk

You, H. S., Kim, K., Black, K., & Min, K. W. (2018). Assessing Science Motivation for College Students: Validation of the Science Motivation Questionnaire II using the Rasch-Andrich Rating Scale Model. *EURASIA Journal of Mathematics, Science and Technology Education*, *14*(4). https://doi.org/10.29333/ejmste/81821

Author Information

Courtney Donovan
https://orcid.org/0000-0001-5911-3294
University of Colorado Denver
United States

Hannah Huvard
https://orcid.org/0000-0003-0799-8814
University of Colorado Denver
United States

Angela Rexwinkle
https://orcid.org/0009-0007-6162-0878
University of Colorado Denver
United States

Robert Talbot
https://orcid.org/0000-0002-6821-0675
University of Colorado Denver
United States
Contact e-mail: *robert.talbot@ucdenver.edu*