Parsons, J. R. M., Parsons, J. C. M., Kohls, K., Sampson, S., Ridolfo, J., (2026). Measuring Foundational Inquiry Skills in First-Year College Composition Courses. *Intersection: A Journal at the Intersection of Assessment and Learning, Early View*.

Measuring Foundational Inquiry Skills in First-Year College Composition Courses

Joshua R. McConnell Parsons, Ph.D., Jannell C. McConnell Parsons, Ph.D., Kathryn Kohls, Ph.D. Shannon Sampson, Ph.D., Jim Ridolfo, Ph.D.

Author Note

Joshua R. McConnell Parsons (0000-0002-0860-4007)
Jannell C. McConnell Parsons (0009-0005-4548-7005)
Kathryn Kohls (0009-0002-8317-1456)
Shannon Sampson (0000-0001-5965-8956)
Jim Ridolfo (0009-0007-5602-4975)

"We have no conflicts of interest to disclose."

Intersection: A Journal at the Intersection of Assessment and Learning
Early View

Abstract: Foundational inquiry skills (FIS) are a critical component of the first-year composition course. Understanding how to navigate resource repositories, assess the quality of information, synthesize source material, and communicate information with stakeholders are critical FIS that ensure future academic success regardless of discipline. Furthermore, these FIS are increasingly critical to students' ability to engage as responsible social citizens both inside and outside the classroom space. Despite the importance of these skills, current tools to measure students' FIS are primarily discipline-focused and/or lack validation. Drawing on multiple years of pre-, post-, and retrospective-pre data, this paper uses Rasch measurement theory to construct and assess the quality of an FIS scale that can be leveraged in inquiry-based first-year college composition classrooms and other entry-point courses focused on FIS.

Keywords: *foundational inquiry, scale validation, Rasch, composition, writing, first-year college students, undergraduates*

Introduction

First-year composition courses are often conceived as the space where college students learn how to write at a college level, with the assumption that writing is a "set of basic, fundamental skills that will apply in other college courses and in business and public spheres after college" (Russell, 1995, as cited in Downs & Wardle, 2007, p. 553). Despite this commonly held conception, writing is highly context dependent, and there is no one right way to communicate that works in every situation. To better prepare students to assess the unique and varied situations they have to communicate in, some first-year composition classes have reoriented to "improve students' understanding of writing, rhetoric, language, and literacy in a course that is

topically oriented to reading and writing as scholarly inquiry and encouraging more realistic understandings of writing" (Downs & Wardle, 2007, p. 553). In other words, this more flexible approach focuses on writing as a process that engages a broad range of foundational inquiry skills (FIS)—including finding and using primary and secondary sources, synthesizing and presenting information, and more. Collectively, these FIS work in concert to help students (1) effectively find, evaluate, and utilize primary and secondary sources as a part of their research process and (2) communicate their findings in nuanced and responsible ways that are adaptive to the needs of the stakeholder audience. Focusing on FIS allows students to flexibly adapt to situational research and communication needs (Council of Writing Program Administrators et al., 2011), preparing them for a wide variety of college research and writing experiences.

While FIS are a critical component of these inquiry-focused first-year writing classrooms, scholarship offers practitioners limited tools for comprehensively measuring FIS in this context. One of the disciplines that we found most active in assessing FIS is library science. Research and college librarians are often tasked with teaching information literacy and student research skills, and the American Library Association has a "Framework for Information Literacy for Higher Education" which includes a "Research as Inquiry" section with learning outcomes that strongly align with our FIS items (Association of College & Research Libraries). Measurement of FIS in library science often takes a more targeted approach, focusing on primary/archival research or secondary research through one-off library instructional days or units (Artman et al., 2010; Hubbard & Lotts, 2013; Leporati et al., 2019) and/or through singular course assignments completed in partnership with the library (Williams & Thayer, 2020). These targeted FIS interventions and their associated assessments are invaluable to first-year composition experience; however, FIS, as outlined in comprehensive first-year composition, also includes critical components of broader contextual communication. The need for broader assessment of FIS that pairs instruction focused on primary/secondary research with the broader outcomes of first-year composition is also recognized within the library sciences community (Artman et al., 2010; Leporati et al., 2019; Williams & Thayer, 2020).

Outside the humanities and social sciences, FIS are frequently assessed in general education STEM literature. In contrast to the more targeted assessments found in the library science literature, this work accounts holistically for a broad range of inquiry-related skills (Familari et al., 2013; Indorf et al., 2019; Weston & Laursen, 2015). These assessment tools are useful for modeling how to study FIS more comprehensively. However, because they are focused on *scientific* literacy and inquiry, many items in these tools lean too heavily on STEM-specific disciplinary knowledge to be useful for assessing FIS in non-STEM classrooms, particularly in first-year composition spaces. While Familari et. al. (2013) do study more general inquiry skill categories such as "inquiry and problem solving" and "communication," for example, they are also concerned with disciplinary skill sets such as "understanding science" and "scientific knowledge" that fall outside of the purview of humanities and social science classroom spaces (p. 7), making some items within their assessment tools less applicable across disciplinary boundaries. Indorf (2019) similarly studies science-specific skills such as "understanding how scientists work on real problems" and "skill in science writing" that, while certainly foundational, are specifically geared toward STEM research and communication (p. 8).

Moreover, while we adapted the general structure and skill areas from Weston & Laursen's Undergraduate Research Student Self-Assessment (URSSA) (2015) to develop our own, we also revised out much of the more science-specific learning outcomes such as "confidence in my ability to contribute to science," "writing scientific reports or papers," ability to "engage in real-world science research," and "feel a part of a scientific community" (p. 5). These studies collectively provide key insights into FIS within a STEM context and curriculum. However, their discipline specificity also falls short of being immediately transferable to other early-career classroom spaces focused on FIS.

Drawing on both these helpful bodies of research on FIS, we developed a 13-item scale that considers FIS more broadly than library studies but is also more geared toward the humanities and social sciences than current STEM models for FIS measurement. While the first-year composition space is the primary focus of the items included in the FIS scale in this study, the scale may be applicable in a broad range of spaces where FIS are the learning outcomes. Although the individual items should provide useful information to scholars and practitioners, determining whether they form a single unidimensional scale measuring the latent construct of foundational inquiry skills can provide a useful metric to support understanding of student development in the field. As a result, the research is guided by the following research question: Do the 13 items included in the Foundational Inquiry Skills (FIS) scale hold together as a single, unidimensional construct when examined using a Rasch measurement approach? To examine this research question, this paper first provides an overview of the development and implementation of the scale, then examines the structural validity and reliability of the 13-item FIS scale and finally considers the potential significance of the availability of such a scale.

Scale Development

The original FIS scale consisted of 13 items loosely adapted from the Undergraduate Research Student Self-Assessment (URSSA) (Weston & Laursen, 2015). Using the URSSA general structure and skill areas, the project team identified key skills emphasized in the inquiry process that are the focus of the first-year composition classroom. Leveraging the project teams' expertise in college composition courses, the team developed the scale items using existing learning outcomes from the home department for the course, as part of the institution's general education learning outcome development. Items on the scale are listed in Table 1. Items 1-6 assess students' confidence in locating, evaluating, and using primary and secondary sources—a critical learning outcome for most first-year writing classes. The remaining items focus on students' confidence in communicating their research to different audiences.

As noted above, college-level writing varies widely based on factors such as audience, situation, and genre, so this set of skills-based items was intended to assess the foundational skills used in almost every college-level writing situation. For each item, students were asked to "please rate your level of skill or ability in the following categories using the accompanying scale." The accompanying scale used a four-point ordinal scale with the response options "no skills," "minimal skills," "average skills," and "advanced skills'' that were transformed into numeric values 1–4 in sequential order (no skill = 1 and advanced skill = 4).

Table 1

Items included on the original FIS scale

Short Code	Full Item
FIS-1	Locating primary sources for research-based projects
FIS-2	Using primary sources in research-based projects
FIS-3	Locating peer-reviewed secondary research for research-based projects
FIS-4	Using peer-reviewed secondary research for research-based projects
FIS-5*	Using [the institution name]'s library system to locate research sources*
FIS-6	Assessing the credibility of a source
FIS-7	Collaborating with others on research-based projects
FIS-8	Adjusting my writing for different audiences
FIS-9	Communicating research to a public audience
FIS-10	Articulating the importance of my research to a public audience
FIS-11	Writing for a research audience
FIS-12	Summarizing source materials for a research audience
FIS-13	Talking about my research interests with my peers

*FIS-5 was removed from the scale after the initial scale validation review.

Methods

Rasch

The Rasch model (Rasch, 1960/1980) is a measurement approach to scale construction that models the probability of a person responding to an item as a function of the difference in a person's ability and an item's difficulty. In survey contexts, this probability reflects the difference between the intensity of a person's latent trait and the degree to which an item on the survey is reflective of that trait, and thus, difficult to endorse. In our case, the latent trait is a student's perceived ability related to FIS, and students respond to a set of items using a rating scale from "no skills" (1) to "advanced skills" (4). Because the response options are polytomous, we used the Rating Scale Model (Andrich, 1978).

The Rasch model assumes that data are unidimensional (Linacre, 1998), placing both items from a test and the individuals taking it along a single continuum of the underlying construct or trait. Winsteps software (version 5.2.2.0, Linacre, 2025g) provides evidence to assess the extent to which the data are indeed unidimensional and that the set of items produces a scale that meaningfully measures the construct of interest.

In ordering items along a single construct, the Rasch model also identifies the hierarchy of item difficulty, providing test users with information on which skills related to the construct are relatively easy to endorse and which are more difficult. Similarly, the Rasch model produces a single-scale measure for people, which can be particularly useful for gauging individual growth

on the latent construct, a common goal when assessing student learning and evaluating educational interventions.

The Rasch model has been utilized as a tool for assessing structural validity of survey instruments across a wide range of social sciences, including library sciences (Scoulas et al., 2020), higher education (Hart et al., 2013), and medical education (Shi et al., 2019). In utilizing the Rasch model for evidence of item functioning and structure of the FIS scale, the research team focused on six criteria that have been utilized widely as metrics for assessment of structural validity: dimensionality, separation and reliability, rating scale effectiveness, item and person fit, and global differential item functioning (DIF) (see Hart et al., 2013)—metrics produced through Winsteps. For a robust description of these metrics for utilization in a validation study, see Scoulas et al. (2020). A summary of these metrics and their associated benchmarks is as follows:

- **Dimensionality:** utilizes a Principal Component Analysis of Residuals (PCAR) to assess if additional latent traits may be present in the data. A target eigenvalue of 2 or less suggests a secondary dimension is unlikely (Linacre, 2025c).
- **Person and Item Separation and Reliability:** indicates the extent to which the scale separates high/lower performers and places items along the latent construct. Benchmark metrics for person separation and reliability are 3 and .8, respectively, and for item separation and reliability are 2 and .9, respectively (Linacre, 2025e).
- **Rating Scale Effectiveness:** provides evidence that categorical response thresholds should sequentially step in alignment with the construct, and thresholds between each categorical measure should advance between 1.4 logits and 5.0 logits (Linacre, 2025d).
- **Item and Person Fit:** indicates how well observed responses conform to model expectations and includes both infit (predictability) and outfit (unexpected behavior). Infit and Outfit for all items are expected to fall between .5 and 1.5 MNSQ (Bond & Fox, 2007). An item with a fit that falls outside the range should be reviewed to determine if it is not clearly written or not well related to the construction.
- **Differential Item Functioning (DIF):** compares either the population examined in the data to a specific subgroup or compares two subgroups together. DIF tests of significance utilize both a Mantel-Haenszel test for dichotomies and a Rasch-Welsh t-test, both of which utilize standard p values ($p < .05$) for evidence of significance in combination with a test of magnitude (above $|X| > .64$ logits for a moderate to large significance) (Linacre, 2025b).

Participants & Procedures

We administered the FIS scale across three academic years to first-year composition students as part of a larger survey that included questions on course outcomes, college academic confidence, and growth mindset (McConnell Parsons et al., 2021). Participants were asked to complete the scale at the beginning (pre) and end (post) of the course.

Participants also reflected on their skills before beginning the course and provided ratings on these reflections as part of the post-survey (retrospective pre). Students were given modest

incentives of additional course points to complete the surveys. Data was collected from two classrooms in year one (pilot) and approximately eighteen classrooms in years two and three. While learning objectives were the same across first-year composition courses, different pedagogical models were employed: one pedagogical model was implemented consistently across all three years as part of a larger examination of that pedagogical model–our treatment group–and two separate comparison pedagogical models were implemented in years two and three–our comparison groups.

Data were collected at a large R1 university in the southeast United States. A total of 446 students provided data across the three years, with 30 from the first year, 166 from the second year, and 250 from the third year. 312 students were from the treatment group and 134 from the comparison groups. Tables 2, 3, and 4 below present the gender, racial, and first-generation status demographics of the study participants, along with the institutional demographics. As the tables show, women may be slightly overrepresented in our data, and Hispanic/Latino students may be slightly underrepresented. Note that percentages listed in Table 2 do not include the individuals who did not provide data.

Table 2
Gender demographics and institutional comparisons

Gender	Participant n	Percentage [a]	Institutional Comparison (undergraduate population)
Male	137	35%	43%
Female	250	65%	57%
Other	1	0%	n/a
Did not provide	58	n/a	n/a

[a] Percentages do not include individuals who did not provide data.

Table 3*Racial demographics and institutional comparisons*

Race/Ethnicity	Participant n	Percentage [a]	Institutional Comparison (undergraduate population)
Asian	22	6%	4%
Black or African American	40	10%	7%
Hispanic or Latino	4	1%	6%
Two or more races	24	6%	4%
Other	5	1%	<1%
White	293	76%	75%
Not provided	58		

[a] Percentages do not include individuals who did not provide data.

Table 4

First-generation status demographics and institutional comparisons

First-Generation Status	Participant n	Percentage [a]	Institutional Comparison (undergraduate population)
Yes	108	26%	28%
No	307	74%	72%
Uncertain	8	n/a	n/a
Did not provide	23	n/a	

[a] Percentages do not include individuals who did not provide data.

To evaluate item functioning and provide structural validity evidence for the scale, we randomly selected one response per participant from the pre, post, and retrospective preconditions to reduce potential bias from local dependency within the data (Linacre, 2025a), ensuring no individual had more than one response in the dataset. For example, when an individual student had responses to all three points in the data set, we used a random number generator to generate a random number between 1 and 3 to determine which time point to retain in the analysis. If a 1 was generated, the pre time point was selected; if a 2 was generated, the retrospective time point was selected; and if a 3 was generated, the post time point was selected for retention. If the student completed only two time points, the number generator randomly selected 1 or 2, with 1 corresponding to the earlier time point in the pairing and 2 to the later. We then entered the data into Winsteps software version 5.2.2.0 (Linacre, 2025e) and assessed the scale's structural validity. For ease of interpretation, Winsteps-produced logit measures were transformed to a 0-100 scale.

Results

During the initial analysis of all 13 items, the fit statistics indicated that the item “Using [institution’s] library system to locate research sources” misfit, with an infit of 1.62 MNSQ and an outfit of 1.74 MNSQ. After reviewing the item conceptually, we determined that this item should be removed because it focused too explicitly on the unique institution rather than broadly on skills applicable to foundational inquiry.

Follow-up analysis revealed evidence of unidimensionality of the 12-item FIS scale. A Principal Component Analysis of Residuals (PCAR) showed that over 50% of the raw variances were explained by measures, and the eigenvalue in the first contrast was 2.59, very close to the target eigenvalue of 2 or less (Linacre, 2025c) (see Table 5). Thus, the data supports the unidimensionality of items.

Table 5

Principal Component Analysis of Residuals (PCAR) of 12-item FIS scale

		Eigenvalue	Observed	Expected
Total raw variance in observations	=	25.1344	100.0%	100.0%
Raw variance explained by measures	=	13.1344	52.3%	51.7%
Raw variance explained by persons	=	10.2902	40.9%	40.5%
Raw unexplained by items	=	2.8442	11.3%	11.2%
Raw unexplained variance (total)	=	12.00	47.7%	48.3%
Unexplained variance in 1st contrast	=	2.5868	10.3%	21.6%
Unexplained variance in 2nd contrast	=	1.6259	6.5%	13.5%
Unexplained variance in 3rd contrast	=	1.2233	4.9%	10.2%
Unexplained variance in 4th contrast	=	1.1683	4.6%	9.7%
Unexplained variance in 5th contrast	=	1.0358	4.1%	8.6%

Next, the scale showed good person-item separation: item reliability was .94, and item separation was 3.87. Person reliability was .89, and person separation was 2.92. These data meet the literature-established benchmarks (Linacre, 2025e), indicating the scale separates high/lower performers and places items along the latent construct. Fit statistics (Table 6) also showed that the items on the FIS scale are productive for measurement, with infit and outfit for all items falling between .5 and 1.5 MNSQ (Bond & Fox, 2007). Analysis of rating-scale categorical response thresholds (Table 7) showed that categorical choices were sequentially stepped in alignment with the construct and that thresholds between each categorical measure ranged from the recommended 1.4 logits to 5.0 logits (Linacre, 2025d).

Table 6

Fit statistics of FIS scale items

Item Short Code	Full Item	Measure in Logits	Measure on 0-100 Scale	Infit MNSQ	Outfit MNSQ
FIS-1	Locating primary sources for research-based projects	0.19	49.06	0.872	0.8773
FIS-2	Using primary sources in research-based projects	-0.25	45.92	0.8846	0.8884
FIS-3	Locating peer-reviewed secondary research for research-based projects	0.65	52.35	1.0129	1.0076
FIS-4	Using peer-reviewed secondary research for research-based projects	0.59	51.90	0.9769	0.9246

FIS-6	Assessing the credibility of a source	-0.15	46.60	1.102	1.0753
FIS-7	Collaborating with others on research-based projects	-0.46	44.41	1.1782	1.2711
FIS-8	Adjusting my writing for different audiences	-0.36	45.12	0.8477	0.8694
FIS-9	Communicating research to a public audience	-0.01	47.62	1.1353	1.1148
FIS-10	Articulating the importance of my research to a public audience	0.04	48.00	0.9413	0.9236
FIS-11	Writing for a research audience	0.51	51.34	1.0108	0.9885
FIS-12	Summarizing source materials for a research audience	-0.13	46.76	0.8999	0.8787
FIS-13	Talking about my research interests with my peers	-0.61	43.32	0.9866	0.9405

Table 7

Andrich Thresholds

Category Label	Observed count	%	Infit MNSQ	Outfit MNSQ	Andrich Threshold	Category Measure
1	215	4	1.09	1.09	NONE	(-4.11)
2	953	18	.93	.92	-2.94	-1.80
3	2819	53	.92	.95	-.65	1.48
4	1338	25	1.06	1.02	3.58	(4.69)

Analyses of differential item function (DIF) also suggest invariance of the scale across racial and gender demographics: no significant differences were observed in either pairwise or global DIF for either demographic group. However, the item “communicating research to a public audience” was significantly more difficult for first-generation students compared to non-first-generation students on a pairwise DIF analysis using the Rasch-Welch t-test for polytomous data (t=-2.28, df=184, p=.0236) with a moderate effect size (.52) (Linacre, 2025b). While a significant DIF existed between the two groups, no significant global DIF was observed, suggesting the scale may not be explicitly biased against the group.

Discussion

This study utilized the Rasch measurement model to examine the structural validity of the FIS scale. After removing the item related to the institute-specific library space, the results provided evidence of unidimensionality and strong item fit, indicating that the items function coherently to measure a single underlying construct. This validity evidence supports the use of the scale as a single summary score, which practitioners and stakeholders can use in practice to assess student progress toward growth in these foundational skill areas.

The inclusion of the library item in the original scale made sense to the team during the development phase and aligned with course learning outcomes; however, the item's misfit on the scale was, in retrospect, unsurprising given the decision to anchor it to the specific institution. Some students who are likely to have confidence in the other items on the scale may not have confidence specifically in the home institution's library system, or vice versa. Furthermore, removing this item did not diminish the scale's quality, as the items related to locating and using primary and secondary sources (FIS-1 through FIS-4) align with the same learning outcomes.

In addition to the misfit, the PCAR revealed three clusters, each loading on two contrasts: items related to locating and using research clustered together, and items related to writing and communication research. This clustering of items, along with the first contrast value exceeding 2.0, could indicate multidimensionality. However, the two contrasts closely align with the learning outcome areas that framed the initial scale development, which we consider essential components of the FIS construct. Moreover, the disattenuated correlation between these two clusters was 0.7594, suggesting that they largely measure the same thing (Linacre, 2025f). As a result, we argue that the clustering within the PCAR matches expectations and supports the scale's unidimensionality.

Limitations

While inquiry spans all disciplines, the foundational skills of navigating sources, assessing the quality of information, synthesizing source material, and communicating that information to stakeholder audiences are often explicitly emphasized in general education composition courses. The FIS scale offers a potential avenue for assessing students' inquiry competencies. However, as with any exploration of structural validity, certain limitations are inherent in the data. While our sample is largely representative of the student population that attended the university where the data were collected, our student population hails from both a research-intensive university and a predominantly white institution.

Future examinations of the validity of the FIS scale may be more robust if the sample were to include more diverse institution types, including two-year colleges, comprehensive universities, and minority-serving institutions. This further exploration is particularly important given that FIS is a key learning outcome of the first-year experience, regardless of institution type. Moreover, we recognize that the potential exists to refine further or expand the scale. While the evidence supports the use of this scale in assessing student learning in the first-year composition course, we recognize that a fundamental trait of scale development is to ensure that the construct's breadth is fully represented. At the same time, we believe we have made a strong effort to include a comprehensive list of FIS taught in the first-year composition course and applicable regardless of the future disciplinary framework students use. Other practitioners may wish to expand the scale to include skills not listed or expand or refine certain items.

Despite these potential limitations, the evidence demonstrates that the 12 items that were retained in the final FIS have a strong case for unidimensionality, and the scale is likely to be useful for documenting change in inquiry skills over time in the first-year composition classroom. Thus, this may be a useful tool for researchers studying innovations in classroom settings where FIS are an outcome of interest.

Implications for Practice

In addition to providing evidence for the structural validity of a scale, Rasch analysis completed in the Winsteps software provides other useful metrics. Primarily, one feature of Rasch measurement is that it places both persons and items along the same continuum of the latent trait—in this case, FIS—and, by doing so, provides a rank order of the difficulty of items for students. Table 8 below presents the logit scores and the transformed scores (on a scale from 0 to 100), with the most difficult items at the top and the easiest at the bottom. As this rank order illustrates, three skills cluster together as generally harder than other skills for students: (R3) locating and (R4) using peer-reviewed secondary research for research-based projects and (R11) writing for a research audience. Conversely (R13), talking about research interests with my peers was the easiest for students. This rank order may be useful in assisting future iterations of first-year inquiry courses as they develop curriculum.

Table 8

Item measures ranked from hardest to endorse to easiest to endorse

Item Short Code	Full Item	Measure in Logits	Measure on 0-100 Scale
FIS-3	Locating peer-reviewed secondary research for research-based projects	0.65	52.35
FIS-4	Using peer-reviewed secondary research for research-based projects	0.59	51.9
FIS-11	Writing for a research audience	0.51	51.34
FIS-1	Locating primary sources for research-based projects	0.19	49.06
FIS-10	Articulating the importance of my research to a public audience	0.04	48
FIS-9	Communicating research to a public audience	-0.01	47.62
FIS-12	Summarizing source materials for a research audience	-0.13	46.76
FIS-6	Assessing the credibility of a source	-0.15	46.6

FIS-2	Using primary sources in research-based projects	-0.25	45.92
FIS-8	Adjusting my writing for different audiences	-0.36	45.12
FIS-7	Collaborating with others on research-based projects	-0.46	44.41
FIS-13	Talking about my research interests with my peers	-0.61	43.32

In addition to the rank order of items, Winsteps also provides a raw-score-to-measure comparison chart that may be useful for practitioners interested in implementing the scale, as shown below for the FIS scale in Figure 1. To use the information in Figure 1, raw numeric scores on each item (see Table 8 for each item)—where a response of "no skills" equals 1, "minimal skills" equals 2, "average skills" equals 3, and "advanced skills" equals 4—are summed then the corresponding raw summed score is applied to this table in the score column. Using this chart, practitioners can convert raw scores to interval-level measures for use in mathematical and statistical analyses, such as assessing students' perceived competencies and tracking individual growth over time. Practitioners can utilize the general 0 to 100 scale to assess growth. However, they can also compare student scores to the item measures (Table 6) to identify where a student falls in relation to the items: a student whose score is below an item measure is unlikely able to endorse their ability in that item confidently—e.g. a student measure score of 46.31 (a raw score of 31) would generally only be confident in their abilities related to the five easiest items on the scale.

Figure 1

Chart converting raw cumulative scores to measure scores

SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.
12	.00E	13.19	25	36.40	3.32	38	63.95	4.37
13	8.99	7.42	26	37.95	3.33	39	66.54	4.23
14	14.52	5.45	27	39.51	3.36	40	68.97	4.11
15	18.02	4.63	28	41.11	3.40	41	71.30	4.06
16	20.71	4.17	29	42.76	3.47	42	73.62	4.08
17	22.97	3.88	30	44.48	3.56	43	76.00	4.18
18	24.97	3.69	31	46.31	3.68	44	78.56	4.40
19	26.80	3.55	32	48.29	3.83	45	81.50	4.80
20	28.52	3.46	33	50.44	4.02	46	85.21	5.57
21	30.16	3.40	34	52.83	4.24	47	90.91	7.49
22	31.76	3.35	35	55.47	4.44	48	100.00E	13.23
23	33.32	3.33	36	58.31	4.54			
24	34.86	3.31	37	61.19	4.50			

Conclusion & Future Research

We recognize the critical importance of developing these foundational inquiry skills to help our students build a strong foundation for interrogating the structural validity and reliability of information as they move through their academic studies and their own personal social and political journeys. Importantly, diverse student populations, including students of color and first-generation students, are more likely to exit college after the first year (Radunzel, 2018), underscoring the need to focus on assessing critical skills, such as inquiry, in the first-year classroom. We intend to examine this FIS scale in other higher education settings (liberal arts colleges, historically Black colleges and universities, community colleges, etc.) to see if the results can be replicated across learning institutions. We hope other researchers will do the same. Furthermore, we hope researchers will explore FIS skills beyond the first year to track transitional skills and skill development as students progress through their majors.

We believe FIS represent a fundamental skill set that prepares students to face the difficult realities of a world where information and truth are often presented in conflicting terms. Engaging students in these skills in their first years of college is critical to their development as responsible social citizens, and therefore, identifying ways for both practitioners and students to understand growth in these areas is essential. This FIS scale, along with the evidence supporting its usefulness, offers one of many tools to help students build the skills needed to navigate both their academic and social spaces.

References

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, *43*(4), 561–573. https://doi.org/10.1007/BF02293814

Artman, M., Frisicaro-Pawlowski, E., & Monge, R. (2010). Not just one shot: Extending the dialogues about information literacy in composition classes. *Composition Studies*, *38*(2), 93–110.

Association of College & Research Libraries. *Framework for information literacy for higher education*. American Library Association (ALA), 11 Jan. 2016, https://www.ala.org/acrl/standards/ilframework.

Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Lawrence Erlbaum Associates.

Council of Writing Program Administrators, National Council of Teachers of English, and the National Writing Project (2011). *Framework for success in postsecondary writing*. https://files.eric.ed.gov/fulltext/ED516360.pdf

Downs, D., & Wardle, E. (2007). Teaching about writing, righting misconceptions: (Re)Envisioning "first-year composition" as "introduction to writing studies. *College Composition and Communication*, *58*(4), 552–584.

Familari, M., Burke Da Silva, K., Rayner, G., Young, J., Cross, A., & Blanksby, T. (2013). Scientific inquiry skills in first year biology: Building on pre-tertiary skills or back to basics? *International Journal of Innovation in Science and Mathematics Education*, *21*(1). http://www.proquest.com/docview/2248389546/abstract/63C03954FD714358PQ/1

Hart, C. O., Mueller, C. E., Royal, K. D., & Jones, M. H. (2013). Achievement goal validation

among African American high school students: CFA and Rasch results. *Journal of Psychoeducational Assessment*, *31*(3), 284–299. https://doi.org/10.1177/0734282912466726

Hubbard, M., & Lotts, M. (2013). Special collections, primary resources, and information literacy pedagogy. *Communications in Information Literacy*, *7*(1), 24–38. https://doi.org/10.15760/comminfolit.2013.7.1.132

Indorf, J. L., Weremijewicz, J., Janos, D. P., & Gaines, M. S. (2019). Adding authenticity to inquiry in a first-year, research-based, biology laboratory course. *CBE Life Sciences Education*, *18*(3), 1–15. https://doi.org/10.1187/cbe.18-07-0126

Leporati, B. R., Bach, P., & Hong, L. (2019). Learning to evaluate sources: Comparing teaching modalities and student outcomes. *Portal: Libraries and the Academy*, *19*(2), 233–252. https://doi.org/10.1353/pla.2019.0014

Linacre, J.M. (1998). Structure in Rasch residuals: Why principal component analysis? *Rasch measurement transactions*, 12(2), 636. https://www.rasch.org/rmt/rmt122m.htm

Linacre, J. M. (2025a). *Dependency and unidimensionality - repeated measures - longitudinal studies*. Winsteps. https://www.winsteps.com/winman/dependency.htm

Linacre, J. M. (2025b). *DIF - DPF - bias - interactions concepts*. Winsteps https://www.winsteps.com/winman/difconcepts.htm

Linacre, J. M. (2025c). *Dimensionality: When is a test multidimensional*? Winsteps. https://www.winsteps.com/winman/dimensionality.htm

Linacre, J. M. (2025d). *Guidelines for rating scales and Andrich thresholds*. Rasch. https://www.rasch.org/rn2.htm

Linacre, J. M. (2025e). *Reliability and separation of measures*. Winsteps. https://www.winsteps.com/winman/reliability.htm

Linacre, J. M. (2025f). *Table 23.1, 23.11, … Principal components/contrast plots of item loading*. WINSTEPS (Version 5.2.2.0). https://www.winsteps.com/winman/table23_1.htm

Linacre, J. M. (2025g). Winsteps (5.2.2.0). Winsteps. https://www.winsteps.com/winsteps.htm

McConnell Parsons, J. R., McConnell Parsons, J. C., Kohls, K., & Ridolfo, J. (2021). Piloting an oral history-based CURE in a general education writing course for first-year students. *Scholarship and Practice of Undergraduate Research*, *4*(2), 27–34. https://doi.org/10.18833/spur/4/2/5

Radunzel, J. (2018). They may be first, but will they last? Retention and transfer behavior of first-generation students. Working Paper 2018–5. Iowa City: ACT.org. Accessed October 22, 2019. http://www.act.org/content/dam/act/unsecured/documents/R1708-retention-firstgen-2018-04.pdf

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests* (expanded ed.), University of Chicago Press.

Russell, D. R. (1995). Activity theory and its implications for writing instruction. In J. Petraglia (Ed.), *Reconceiving writing, rethinking writing instruction* (pp.51–78. Routledge. https://doi.org/10.4324/9780203811948

Scoulas, J. M., Aksu Dunya, B., & De Groote, S. L. (2021). Validating students' library experience survey using the Rasch model. *Library & Information Science Research*, *43*(1). https://doi.org/10.1016/j.lisr.2021.101071

Shi, Y., Gugiu, P. C., Crowe, R. P., & Way, D. P. (2019). A Rasch analysis validation of the Maslach

Burnout Inventory—Student survey with preclinical medical students. *Teaching and Learning in Medicine*, *31*(2), 154–169. https://doi.org/10.1080/10401334.2018.1523010

Weston, T. J., & Laursen, S. L. (2015). The Undergraduate Research Student Self-Assessment (URSSA): Validation for use in program evaluation. *CBE Life Sciences Education*, *14*(3). https://doi.org/10.1187/cbe.14-11-0206

Williams, C. B., & Thayer, J. (2021). Connecting community college students to primary sources in the archives. *College & Undergraduate Libraries*, *28*(1), 119–127. https://doi.org/10.1080/10691316.2020.1739585

About the Authors

Joshua R. McConnell Parsons, Georgia Institute of Technology, joshua.parsons@gatech.edu
Jannell C. McConnell Parsons, Georgia Institute of Technology, jparsons60@gatech.edu
Kathryn Kohls, University of Kentucky, kathryn.kohls@uky.edu
Shannon Sampson, University of Kentucky, shannon.sampson@uky.edu
Jim Ridolfo, University of Kentucky, ridolfo@uky.edu