Foundational Competencies in Educational Measurement: A Rejoinder

Share Your Story

Foundational Competencies in Educational Measurement: A Rejoinder

Task Force Members (Authors):

Andrew D. Ho*, *Harvard Graduate School of Education*
Terry A. Ackerman, *University of Iowa*
Deborah L. Bandalos, *James Madison University*
Derek C. Briggs, *University of Colorado Boulder*
Howard T. Everson, City University of New York
Susan M. Lottridge, *Cambium Assessment, Inc.*
Matthew J. Madison, *University of Georgia*
Sandip Sinharay, *Educational Testing Service*
Michael C. Rodriguez, *University of Minnesota*
Michael Russell, *Boston College*
Alina A. von Davier, *Duolingo*
Stefanie A. Wind, *University of Alabama*

* Corresponding Author:
Andrew Ho
455 Gutman Library
6 Appian Way
Cambridge, MA 02138
andrew_ho@gse.harvard.edu

Foundational Competencies in Educational Measurement: A Rejoinder

What are foundational competencies in educational measurement? We published a framework for these foundational competencies in this journal (Ackerman et al., 2024) and were grateful to receive eight commentaries raising a number of important questions about the framework and its implications. We identified five cross-cutting recommendations among the eight commentaries relating to 1) our process and purpose, 2) Artificial Intelligence, 3) ethical competencies, 4) qualitative, critical, and culturally responsive commentaries, and 5) intersecting professions in, for example, classroom assessment and content development. In this rejoinder, we respond to these five recommendations and to each of the eight commentaries in turn. We hope that discussion and consensus about foundational competencies in educational measurement continue to advance in our journals and our field.

In October 2021, then President of the National Council on Measurement in Education (NCME) Derek Briggs formed a Task Force on Foundational Competencies in Educational Measurement and charged it to:

1. develop a set of foundational competencies for the field of educational measurement,

2. illustrate one or more curricular models for a graduate program in educational measurement, and

3. engage NCME membership and the field with Task Force findings through conference presentations and published journal articles.

The Task Force released a draft report for NCME member comment in 2022; engaged the NCME community in open webinars; published a full report on the NCME website in March 2023 (Ho et al.,

2023); hosted symposia with commentary from NCME members at the 2023 annual meeting, and published a condensed, peer-reviewed, and edited version of the report in this journal (FCEM, *Foundational Competencies in Educational Measurement*, Ackerman et al., 2024).

We are grateful to editor Zhongmin Cui for selecting this article as the focus of an open call for commentaries. We thank the authors of the 8 published commentaries for engaging with our work and inspiring us with their ideas. The 12 members of the FCEM Task Force read the commentaries and met to discuss and develop this consensus rejoinder. Here, we describe and discuss five cross-cutting recommendations that we identified as emergent from many of the eight commentaries. We then proceed to respond briefly to each of the eight commentaries in turn.

The five cross-cutting recommendations that the Task Force identified among the eight commentaries are to:

1. Increase clarity around the process, purpose, and positionality of the Task Force.
2. Emphasize Artificial Intelligence in the FCEM framework
3. Add ethical competencies to the FCEM framework
4. Add complementary (e.g., qualitative, critical, culturally responsive) competencies to the FCEM framework
5. Broaden the framework to encompass intersecting professions (e.g., classroom assessment professionals, test security professionals, and content development professionals).

Recommendation 1: Increase clarity around the process, purpose, and positionality of the Task Force.

We appreciate calls from some of the commentaries to clarify the process, purpose, and positionality of the FCEM Task Force. For example, Araneda et al. (2024) ask us to clarify our intentions, "What we describe as a consensus document for the foundational competencies in our field will shape that future. How intentional was the Task force in wanting to shape this future?" (p. xx). Crespo Cruz et

al. (2024) state similarly, "it would be helpful to understand further the motivation for the report and the ways in which this information should and could be used" (p. xx). And van Orman et al. (2024) observe, "the team lacked representation from emerging scholars and graduate students."

As a peer-reviewed journal article, our focal article followed the norms of quantitative measurement research. As a general matter, articles in this journal do not include many details on the construction of the research team, the review and consensus process, or positionality statements. These norms are changing and certainly deserve further scrutiny (Castillo & Gillborn, 2023). This rejoinder provides an opportunity to re-establish some of the context around process, purpose, and positionality that is available in the full Task Force Report (Ho et al., 2023). In that report, Past President Derek Briggs used a foreword to reflect upon the many functions such a report could serve:

> Such a document could help to improve the visibility and standing of the field. It could attract talented and committed undergraduate students to important work. It could cohere and improve instruction among and within graduate programs. It could improve the skills and readiness of incoming professionals to measurement organizations. And it could even serve as the basis for a license and certification program in educational measurement (Ho et al., p. 4).

Past President Briggs also described his process for identifying Task Force members in his foreword:

> I asked for and received nominations from NCME's Board of Directors and from the co-chairs of NCME's Educators of Measurement (EoM) Special Interest Group in Measurement in Education (SIGIMIE). From these nominations, I identified 11 individuals who were diverse in terms of their role in the field of educational measurement, their areas of expertise, their years of experience, and their gender and racial/ethnic identity (Ho et al., p. 4).

The full report also documents the outreach process meant to fulfill the third part of the charge that Past President Briggs imparted to the Task Force, "…to engage NCME membership and the field…" (Ho et al., 2023). All NCME members received an invitation to comment on a draft version of the Task Force report in September of 2022, and we held an open webinar to discuss the draft framework. We received 17 comments and posted them all on the Task Force website with our responses to each of them (Ackerman et al., 2023). Three of these comment authors are also authors in current commentaries (Brookhart, 2024; Araneda et al., 2024; Crespo Cruz et al., 2024). The Task Force also hosted two symposia at the 2023 NCME annual meeting inviting external commentary (NCME, 2023). Two additional commentaries in this collection had their roots in those symposia (Cheng et al., 2024; Middlebrook et al., 2024).

We also observe that these commentaries and this rejoinder are themselves part of a participatory process. The outreach has continued and will continue. As we hear observations like Araneda et al. (2024), that, "this focal document, its construction procedure, is continuing a different view about measurement and politics, in which science is protected in a kind of 'ivory tower,' isolated from external influences and independent from moral judgments" (p. xx), we understand that continued facilitation of dialogue with NCME members and the field will be necessary to build consensus. As Task Force chair Andrew Ho stated in his afterword of the full Task Force Report (Ho et al., 2023):

> I know our task force members are committed to continued effort toward consensus about foundational competencies in educational measurement. Discussion and debate over foundational competencies can improve coherence in the field and the NCME professional community. As our report stated in our 6 principles at its outset, we expect foundational competencies to shift as society and science advance. We welcome continued engagement and periodic revisitation and advancement of this document by others.

Recommendation 2: Emphasize Artificial Intelligence in the FCEM framework

We also appreciate the recommendations of many commentary authors to elevate consideration of Artificial Intelligence (AI), including natural language processing (NLP), machine learning (ML), and large language models (LLM), in our framework. Crespo Cruz et al. (2024) noted, "the foundational competencies mention AI and NLP but make no effort to define the scope or content of these terms" (p. xx). Rupp (2024) suggests that "advances in AI require educational measurement specialists to re-examine the role of 'humans in the loop' during the design, implementation, evaluation, and refinement of [measurement] technologies" (p. xx). And in their review of job postings related to psychometrics, Cheng et al. (2024) note that AI-related terms account for 4.5% of all keywords and conclude that "ML, NLP, and LLMs have a visible influence on educational measurement" (p. xx).

We had many discussions about Artificial Intelligence on our task force that includes two members, Susan Lottridge and Alina von Davier who have considerable expertise and experience, in the areas of AI and Computational Psychometrics (e.g., Hao et al., 2024; von Davier et al., 2022). Two other task force members also authored recent commentaries on AI and educational measurement (Briggs, 2024; Ho, 2024) motivated by emerging standards in the field (e.g., Burstein, 2024; International Test Commission & Association of Test Publishers, 2022). Our Task Force currently highlights the importance of AI in three areas, Domain 2 (Technical, Statistical, and Computational Competencies), Domain 3B (Theory and Instrumentation), and Domain 3E (Psychometric Modeling). Do AI competencies deserve further elevation? For example, should we rename Domain 2 to be Technical, Statistical, and AI Competencies?

AI encompasses a diverse and evolving set of technologies that are rapidly changing the way we live and work. Rather than attempt to define AI restrictively or predict its proliferation, our Task Force discussions centered on how AI tools can, will, and do serve educational measurement. We were not charged with identifying Foundational Competencies in Artificial Intelligence. And we do not believe that

educational measurement scholars and professionals must have foundational competencies in the design and improvement of specific AI tools like LLMs and NLP algorithms (though some may). The "foundational competencies" we identified that related to AI were therefore statistical, computational, and psychometric competencies that underlie the use and development of AI tools for educational measurement. We identified "theory and instrumentation" as a particularly promising competency subdomain where AI and computational competencies could help with item generation, construct refinement, and task design. The field's uses of AI in measurement theory and instrumentation appear to be advancing rapidly, and we encourage ongoing consideration of the relevant foundational competencies, as well as collaboration (Domain 1) with those with advanced computational competencies (Domain 2), in the face of this change.

We therefore believe that AI is already addressed in our framework both in terms of the relevant foundational competencies underlying AI and the ways in which AI can advance educational measurement efforts. We see "AI" as a diverse and versatile set of tools, not a foundational competency in educational measurement itself. As our field and other fields rightly elevate the importance of the "human in the loop" of AI systems in education, it is our belief that the most important humans to put in these loops are educational measurement experts, trained to elevate validity and fairness (Subdomain 3B) in context (Subdomain 3A), in particular.

Recommendation 3: Add ethical competencies to the FCEM framework

Many commentary authors suggested that we add ethical competencies to our FCEM framework. Araneda et al. (2024) state, "the ethical dimension should be foundational. By not including it, we can create the impression in the document that an amoral type of practice, in which normative aspects of using testing as a technique are disregarded, is acceptable" (p. xx). Middlebrook et al. (2024) similarly argue, "as society evolves, the ability to view the world, including measurement situations, through ethical lens, becomes more critical" (p. xx). And van Orman et al. (2024) suggest, "it is necessary to

foster competencies that encompass both technical abilities and a thorough understanding of the ethical and social implications of these technologies to avoid inadvertently perpetuating biases or inequities" (p. xx).

The Task Force agrees that understanding ethical dimensions of test use and score interpretation are essential for educational measurement professionals in specific roles. However, we had trouble identifying specific ethical competencies as "foundational" that are not already addressed by principles of fairness (Subdomain 3B) in social, cultural, historical, and political contexts (Subdomain 3A). In doctoral programs, students learn about ethical principles and practices in the context of conducting human subjects research given the primary roles and responsibilities of doctoral students as researchers. Students with internships or assistantships in measurement organizations should also have the opportunity to learn about ethical professional practices. These specific ethical principles tend to be role-specific, and thus they were difficult to describe as foundational. The *Standards for Educational and Psychological Testing* (AERA et al., 2014) similarly defer ethical considerations to relevant professional contexts and organizations.

Decades ago, NCME formed an "Ad Hoc Committee on the Development of a Code of Ethics." They ultimately published a "Code of Professional Responsibilities in Educational Measurement" (Schmeiser et al., 1995), with responsibilities separated by 8 different roles (e.g., those who develop assessments, market assessments, purchase assessments, administer assessments, and so on). The necessity they saw of delineating ethical principles by role is consistent to us with our argument that the foundational competencies supporting ethical principles and behavior are related to fairness in context rather than ethical educational measurement competencies in general. The NCME board has recently discussed revisiting and potentially revising the 1995 code. The Task Force welcomes such a complementary effort. Meantime, we believe that identifying and adhering to ethical principles can be guided by developing foundational competencies in fairness and context (Subdomains 3A and 3B).

Recommendation 4: Add complementary (e.g., qualitative, critical, culturally responsive) competencies to the FCEM framework

Various commentary authors wished for us to identify a broader set of competencies. Araneda et al. (2024) claim that, "the document does not include anything related to qualitative research" (p. xx). Crespo Cruz et al. (2024) wish for more "depth and specificity" in Subdomain 3A related to context, arguing that, "there is an ongoing need for critical reevaluation in the field regarding fairness, race, and culture, with increasing focus on 'cultural sensitivity' and 'cultural responsiveness'" (p. xx). Van Orman et al. (2024) state, "as the field continues to embrace a more collaborative ethos underscored by its evolving interdisciplinary nature, the significance of integrating a culturally responsive lens into these competencies cannot be overstated" (p. xx).

The Task Force appreciates this encouragement to push our competencies beyond those that are narrowly quantitative and psychometric. From our first meetings, Task Force members shared our mutual appreciation that measurement is not purely quantitative but requires theory, communication, collaboration, instrumentation, and context to master. In their encouragement, commentary authors acknowledged that we dedicated a domain to Communication and Collaboration (Domain 1); elevated the importance of social, cultural, and historical context (Subdomain 3A); and identified theory as a foundational competency (Subdomain 3C). We were not convinced in our reading of these encouragements that our framework required fundamental restructuring or renaming, but we welcome the elaborations, references, and specificity that these authors suggested as they interpret, apply, and extend our framework.

One perspective that the Task Force discussed was encapsulated in a mantra that Task Force Chair Andrew Ho teaches his students: "Measurement must be qualitative, then quantitative, then qualitative again." This means that measurement must begin with questions of what we measure and why, qualitatively and theoretically, in a context and in a community of practice (Domain 1, Subdomains

3A and 3C). Measurement must then become quantitative using instrumentation and statistical and psychometric methods (Domain 2, Subdomains 3C-3E). Then, these quantities must again become qualitative: communicated, interpreted, and used in context (Domain 1, Subdomains 3A and 3B).

In these ways, we believe that we are elevating qualitative, critical, contextual, culturally responsive, and collaborative competencies, even if commentary authors may wish for more emphasis on these competencies or different terms to describe them. Rather than pitting the quantitative against the qualitative or other complementary competencies in opposition, we hope our framework highlights how these competencies in their intersections and overlap are necessary for fair and valid educational measurement practices.

Recommendation 5: Broaden the framework to encompass intersecting professions (e.g., classroom assessment professionals, test security professionals, and content development professionals).

Authors of three commentaries advocated strongly for the inclusion of competencies for non-psychometric professionals who are part of the educational measurement community.
Wine & Hoffman (2024) contend that our framework, "insufficiently considers the vital contributions of content development professionals [CDPs] to the real work of educational measurement by omitting the CDP perspective from any of these domains or subdomains" (p. xx). Brookhart (2024) states that she "[does] not see foundational competencies for classroom assessment careers in the [FCEM]," and she proposes that our FCEM should be retitled as, "foundational competencies for careers in large-scale educational testing" (p. xx). And Araneda et al. (2024) state that, "the foundational competencies do not include test security" (p. xx). They suggest including test security issues as examples in our discussion of fairness (Subdomain 3B) and theory and instrumentation (Subdomain 3C).

We return to the six principles we stated at the outset of our article, particularly principles 1 and 2: foundational competencies are not an exhaustive list of competencies, and they also overlap and interact with competencies in other professions and disciplines (Ackerman et al., 2024). Content

development professionals, classroom assessment professionals, and test security professionals all have their own competencies that intersect with our FCEM framework. We agree with these authors that FCEMs do not always align exactly with the needs of these or other professions. This is understandable, because these professions require not only foundational competencies but advanced and specialized competencies, and because these professions intersect with but do not require what we consider to be the full set of competencies that are foundational for educational measurement.

Because some authors have associated our competencies with large-scale testing or psychometrics, we should be more explicit about our position. As Briggs (2022) made clear in his NCME presidential address: "testing and measurement are not the same thing" (p. 401). Additionally, as our definitions in Figure 1 (Ackerman et al., 2024) make clear, *psychometrics and measurement are not the same thing*. Indeed, some of our own Task Force members who have expertise in psychometrics do not feel that they have full foundational competencies in, for example, subdomain 3A (Context) and 3C (Theory and Instrumentation). This is why we elevated Principle 4 in our article: "foundational competencies are both descriptive of the profession and discipline and aspirational about the future." We believe this principle applies to us as well as the field.

In the arguments of our colleagues, we hear that we could have been even more aspirational, as well as clearer and more specific about some of these competencies related to context, theory, and instrumentation. We also note that these commentaries came from scholars with relative expertise in qualitative methods, classroom contexts, and substantive content. We believe that this expertise has too often received short shrift in NCME, an organization that has disproportionately privileged competencies in statistical and psychometric domains. We believe our FCEM framework is clear about valuing additional competencies, and we look forward to working with our colleagues to further elaborate upon additional competency domains.

Although we are an NCME Task Force, we should also be clear about our intentions with our efforts in developing our framework: Any attempt to define foundational competencies in educational measurement should not circumscribe the competencies that NCME as an organization should value. These competencies include those of classroom assessment professionals, test security professionals, content development professionals, and, yes, psychometric professionals. All of these professional competency domains extend beyond our FCEMs, and none fully subsumes our FCEMs. This is why we believe our framework is a step toward a coherent and aspirational understanding of the "foundational competencies in educational measurement."

We proceed here to address each of the 8 commentaries briefly in turn, in alphabetical order by the first author's last name.

Araneda, Fremer, & Foster (2024)

These authors were critical of our report, stating that a student trained according to our FCEM framework, "will not be fully capable of participating in the testing industry…" (p. xx). They took issue with a number of procedural and substantive decisions that the Task Force made, some of which we address above in our sections on process, AI, ethics, and test security. We address a few additional issues that the authors raise here.

First, many competencies that Araneda et al. (2024) felt were missing or neglected were those for which we felt we had in fact identified foundations. Rather than list a series of topics (test security, artificial intelligence, legal aspects, ethics, qualitative research), the Task Force's charge was to identify whether there were foundational competencies underlying these topics. As we argue in Recommendation 2, rather than "teach AI," we believe learners should develop computational competencies (Domain 2) and deploy these in theory and instrumentation (Subdomain 3C) and psychometric modeling (Domain 3E). Rather than "teach ethics," we believe that learners should develop competencies related to fairness (Subdomain 3B) in context (subdomain 3A). The authors

themselves reflect that Test Security is a fundamental fairness issue (subdomain 3B) that can be addressed in part in theory and instrumentation (subdomain 3C). We agree, and in this way, we believe that the FCEM framework identifies foundational competencies needed for a broad range of educational measurement endeavors, including test security.

Second, we appreciate the authors' critique of the document as seeming antidemocratic, apolitical, and insulated. We hope our discussion under Recommendation 1 above adds additional context. We also feel that the authors could have reflected further that the document does not exist on its own as if chiseled in stone, and that their commentary, this rejoinder, and our assured ongoing discussion of this topic in the future is part of the participatory process that was an explicit element of our charge.

Finally, the authors state that, "this document did not have any consultation with graduate students at NCME" (p. xx). As we described in our response under Recommendation 1, above, all NCME members, including graduate student members, were invited to comment on the draft report. We observe that the Araneda et al. commentary had its roots in a graduate student comment on this draft report, and the Task Force responded to this comment when we posted our final report in March 2023 (Ackerman et al., 2023). In addition to a webinar and listening session for all NCME members on October 5, 2022, we held a webinar for the Educators of Measurement special interest group on October 11, 2022, where we encouraged educators to engage their students around the FCEM draft framework. We appreciate the encouragement to consult further with graduate students, and we welcome feedback either formally through the NCME Graduate Student Issues Committee or through other channels.

Brookhart (2024)

Brookhart (2024) identifies misalignment between the foundational competencies in educational measurement that we present and the competencies necessary for classroom assessment. She recommends renaming the document to clarify that we identify competencies for careers in large-

scale educational testing. To the Task Force, classroom assessment intersects with educational measurement, but the two practices are different and, as Brookhart observes, require different competencies.

As we define our terms in Figure 1 of our paper (Ackerman et al., 2024), assessment does not require measurement (see also Briggs, 2022). Therefore, we do not believe that educational measurement is, as Brookhart suggests, "all-encompassing," and it certainly does not and should not encompass classroom assessment. We welcome the development of foundational competencies in classroom assessment, and much of Brookhart and others' work (e.g., Klinger et al., 2015) provides a basis for such development.

Brookhart rightly notes that classroom assessment has been a high-priority focus of NCME for years. Classrooms are an important context (Subdomain 3A) within which teachers and students may use educational measurements as well as other assessment procedures. Brookhart's argument suggests that some could misread our effort as narrowing the scope of the competencies that NCME, as an organization, values and upholds.

As we note in our response to Recommendation 5 above, our charge was to identify foundational competencies in educational measurement, and these competencies are not exhaustive and intersect with other competencies. As NCME members, we value the diversity of competencies that we and our fellow members bring beyond educational measurement. In September of 2024, there will be a symposium at the NCME Special Conference on Classroom Assessment entitled "Comparing foundational competencies in educational measurement and classroom assessment." That this symposium and conference would be held under the NCME banner reflects the way in which educational measurement competencies support many intersecting professions, all of whose professionals are welcome at NCME.

Cheng, Pei, Filonczuk, & Le (2024)

Cheng et al. conducted a text analysis of 80 job postings related to psychometrics from July 1, 2023, to January 25, 2024. They categorize 562 keywords into six mutually exclusive categories and calculate percentages by frequency pooled across postings: soft skills (23%), statistics (15%), measurement (23%), research capabilities (18%), programming and software (17%), and AI (5%). We appreciated this quantitative approach to analyzing job postings and agree with the authors' conclusion that, "in general, keywords extracted from the job posts map well to the foundational competencies" (Cheng et al., p. xx). We have three additional comments on their analysis.

First, the Task Force wondered how the analysis might have been different had the initial search been broader than variants of the term "psychometrician." As part of our goal was to broaden educational measurement to encompass other competencies beyond the psychometric, we wondered whether a search that included "measurement," "testing," "assessment," and combinations of other terms might have yielded different results, perhaps with more skills related to FCEM Subdomain 3A that Cheng et al. (2024) observe was not represented. Part of our hope is that our FCEM framework leads to a better understanding of "educational measurement specialist" as a professional role.

Second, Cheng et al.'s categorization of skills into mutually exclusive categories is at odds with our FCEM framework, which envisions intersections and overlap among domains and subdomains. We could collapse their 6 categories into our three domains, yielding 23% in Domain 1 (soft skills = Communication and Collaboration), 55% in Domain 2 (statistics + research + programming + AI = Technical, Statistical, and Computational Competencies), and 23% in Domain 3 (measurement = Educational Measurement Competencies). However, we wonder whether an alternative classification system that allows for terms to belong to multiple domains and subdomains would yield contrasting conclusions.

Third, we appreciated Cheng et al.'s transparency in enabling their analysis to be replicated and extended. We agree with their suggestion that repeating this analysis at regular intervals would be useful to NCME as well as sibling organizations related to psychometrics, testing, and measurement. If or as we do so, perhaps we will see what we hope, that foundational competencies in educational measurement can rise in prominence and desirability across a wide range of professions.

Crespo Cruz, Immanuel, Keller, Ketan, McIntee, Mena Serrano, Sireci, Smith, Suárez-Álvarez, Wells, Woodland, & Zenisky (2024)

We were heartened to see a joint response from faculty and students at an established measurement program, in this case the University of Massachusetts Amherst. We hope their enthusiastic and critical response is a model for other measurement program faculty and students to engage with the FCEM framework. Some of the issues that Crespo Cruz et al. (2024) raise are addressed in our responses above, including our elaboration of our process and purpose (Recommendation 1), how we address AI (Recommendation 2), and how we address qualitative and critical competencies (Recommendation 4). We add a few additional responses here.

At one point of their commentary, Crespo Cruz et al. (2024) misread our framework, suggesting that, Subdomains A and B (which they mislabel as A and E) "are positioned at the margins of theory, precision, and modeling, almost as sidenotes, rather than as foundational to everything we do" (p. xx). As we state in the text, "Subdomain A is overarching competency subdomain" (p. xx), and "Subdomain B is an undergirding competency subdomain" (p. xx). We also state in Principle 3 that domains "overlap and interact with each other" (p. xx). Such a profound misreading of our key figure suggests that we must be clearer in our presentations to align the figure with our text and our intentions. Our article also includes an alternative Figure A1 in Appendix A that may be clearer in its emphasis of the overarching, undergirding, and intersecting nature of our FCEM domains and subdomains (Ackerman et al., 2024).

At another point, Crespo Cruz et al. (2024) misread our description of a "traditional course sequence" as advocacy. We stated, "traditional course sequences in educational measurement often begin with a treatment of validity and defer fairness and methods for detecting differential item or test functioning until later in curricular sequences" (Ackerman et al., 2024, p. xx). Crespo Cruz et al. object to this: "This reflects our general sense that in a traditional approach, some topics such as equity and fairness are generally marginalized and pushed to later in the test development process, reflecting a bias to maintain historical or traditional processes that bake systemic inequality into our work and our training of the next generation of measurement professionals" (Crespo Cruz et al., 2024, p. xx). In fact, their reaction is entirely consistent with the next sentence in our article, where we say, "in contrast, developing validity, validation, and fairness as an undergirding foundational competency requires elevating these concepts such that they are visible in all educational measurement activities throughout the curriculum" (p. xx). As the turn of phrase goes, Crespo Cruz et al. (2024) and the Task Force are "in vehement agreement" on this point, and we will endeavor to be clearer about this.

Finally, we appreciate Crespo Cruz et al.'s attention to trends in international educational measurement work as well as international student representation in U.S. graduate programs. We believe that foundational competencies should apply to international educational measurement efforts with particular attention to competencies from Subdomain 3A (Context) and Domain 1 (Communication and Collaboration). As Zumbo (2014) notes in his review of the 2014 Joint Standards (AERA et al., 2014), fundamental concepts can transfer across borders. However, the degree to which foundational competencies cross borders warrants further discussion with international organizations and communities. Such outreach is ongoing, with symposia related to this FCEM framework planned at the International Test Commission and Psychometric Society annual meetings in the summer of 2024.

Middlebrook, Hamilton, & Walker (2024)

We appreciated this commentary from three scholars at different career stages calling for shared responsibility among assessment and research organizations, doctoral training programs, and others among the community to support students and early career professionals toward lifelong learning. We support in particular their recommendation, "for employers to **create the conditions and learning opportunities to promote on-the-job development of a broad set of competencies**" (Middlebrook et al., p. xx, emphasis in original). The authors also encourage us towards greater depth and specificity in many of our competency domains, including advocacy in communication and collaboration (Domain 1) and real data applications in technical, statistical, and computational competencies (Domain 2).

In Ackerman et al. (2024), we acknowledged the importance of a number of dispositions in Principle 6, including intellectual humility, flexibility, and creativity, but distinguished these from our competencies. We nonetheless appreciate the efforts of Middlebrook et al. (2024) to describe dispositions they felt were important, including active listening, inclusivity, and thoughtful skepticism. It is possible that the line between dispositions and what we identified as communication and collaboration competencies (Domain 1) may be blurry. We encourage greater consensus about these dispositions or collaborative competencies, and we hope we can explore how we can develop them in the context of educational measurement.

We also appreciated Middlebrook et al. (2024) challenging our definition of a career in educational measurement as supporting measures of "cognitive, affective, and other psychological constructs" (Ackerman et al., p. xx). Middlebrook et al. argue that "an emphasis on psychological constructs misses opportunities to prepare measurement professionals at the group or institution level" (p. xx). We agree with this, and we further believe that measurement competencies can extend not simply to aggregated scores of psychological constructs, but also to measurement of groups and institutions themselves (NASEM, 2019; Raudenbush & Sampson, 1999). These are the contexts in which

psychological constructs change (Subdomain 3A), and they are therefore necessary to consider and even incorporate formally into psychometric models (Borsboom, 2022).

Rupp (2024)

Rupp (2024) offers important perspectives on the necessity of agile innovation in the professions. It was indeed this current period of rapid change in technology and education that made our Task Force charge such an interesting challenge for us: Are there foundational competencies that prepare us enduring careers in educational measurement in the face of constant change? Although our FCEM framework is an answer, Rupp wonders, "whether the value proposition for a graduate degree in educational measurement can remain strong enough for the job market of the future as it will have to compete with the value proposition offered by similar kinds of technical and computational degrees from neighboring disciplines" (p. xx).

At the same time, Rupp (2024) observes, "a notable disconnect between the high degree of technical specialization in advanced modeling commonly afforded by a 'prototypical' Ph.D. education with a methodological focus and the much higher degree of reliance on a wide array of applied, interdisciplinary skill sets in real-life jobs" (p. xx). It is exactly this observation that encouraged us toward an FCEM framework that extends beyond the technical and psychometric skills that are the relative strength of some measurement programs. Rupp wonders, provocatively, whether such skills could be achieved through "a well-targeted Master's degree coupled with diverse job experiences and strong qualifications in so-called 'skills for the future'" (p. xx). We welcome the possibility of broader training via modules and Master's degrees, but we also hope our FCEM framework inspires doctoral training programs in educational measurement to either find their competencies in the framework or fill in competencies that may be currently overweighted toward the technical and psychometric.

Van Orman, Jackson, Vo, & Taylor (2024)

Van Orman et al. (2024) offer provocative observations about our process and positionality and go further to propose an alternative definition of measurement and an alternative graphical framework. We are grateful for their engagement and intrigued by their alternative conceptions. We addressed some of their questions and critiques related to representation (Recommendation 1), ethics (Recommendation 3), and qualitative methods (Recommendation 4) earlier in this rejoinder.

Van Orman et al. (2024) adapt our definition of measurement to add, "instrumentation that results in **[qualities and/or]** quantit**[ies]** intended to support inferences" (Ackerman et al., 2024, in plain text, Van Orman et al., 2024, in bold). We believe this definition is closer to our definition of "assessment," and we do not believe qualities alone can be a measure. Qualities certainly hold value. They serve many purposes, including, as we note in our response to Recommendation 4, playing an essential role in guiding theory and instrumentation (Subdomain 3C) and, in qualitative score interpretations, serving as the target for validation (Subdomain 3B). As we emphasize in our response to Recommendation 5, we do not believe our framework excludes any non-quantitative assessment efforts or experts from NCME, just as it would not exclude a narrowly quantitative psychometric effort or expert lacking in non-quantitative measurement competencies.

Van Orman et al. (2024) also argue that educators, education leaders, and policymakers should be included as "measurement professionals." We certainly welcome people in these roles to develop our FCEM competencies, and we understand educational measurement professionals must engage with people in these roles to ensure valid test score interpretations and uses. However, we believe that our full framework of FCEM competencies are not necessary for those in these roles. We believe it remains useful to distinguish those who support the design, use, and evaluation of measures (educational measurement professionals) from those who use measures (educational measurement users).

Finally, there is a lot we find appealing about the framework for foundational measurement skills that Van Orman et al. (2024) propose. We appreciate the framework's orientation toward growth

through the metaphor of a plant, its emphasis on the interactions among competencies via roots, and the distinct roles of the sun, soil, and water. The authors did not provide full explanations for all labels and locations of concepts in their diagram, so we encourage them to develop their framework further, in line with their definition of measurement and their goals. And we appreciate the feedback that our figure on its own may leave a reader with the impression Subdomains 3A and 3B are siloed, as this was a misunderstanding in other commentaries.

Wine & Hoffman (2024)

We appreciate the reminders by Wine and Hoffman (2024) that content development professionals (CDPs) are essential to the "large scale, on-demand, criterion-referenced standardized assessment" (p. xx) process. We understand that the authors felt our extensions beyond psychometric modeling were insufficient in our characterization of these additional competencies. We welcome their attempts to provide references and language to elaborate on the importance of theory, instrumentation, collaboration, and communication. Although we address the central issue that Wine and Hoffman raise in our response to Recommendation 5 above, we address a few additional points here.

The authors argue that "respect for colleagues" and "listening" should be elevated as competencies. We welcome these as collaborative competencies. The authors claim that we, "present communication as just making oneself understood, to the seeming exclusion of trying to hear and working to understand others" (Wine & Hoffman, p. xx). We do not feel that this is an accurate characterization of our full description of Domain 1. We stated, "collaboration competencies include the skills required to work in a constructive manner with other professionals… co-creating solutions in ways that synthesize or build on ideas… to understand a variety of perspectives" (Ackerman et al., 2024, p. xx). While we do not state "respect" nor "listening" explicitly, it is difficult for us to imagine achieving what we describe without these competencies.

Like Crespo Cruz et al. (2024), Wine and Hoffman (2024) appear to have misread our figure as suggesting that Subdomain 3B (validity, validation, and fairness) is separate from other subdomains like Theory and Instrumentation. Wine and Hoffman (2024) state, "validity and fairness are woven in with the work of item and instrument development (Ebel, 1951), which are not so much as mentioned in this section" (p. xx). Because Wine and Hoffman appear to neglect our framework's earlier emphasis on intersections, they do not appear to count our statement in Subdomain 3C, where we state clearly, "sound instrument design and development is a critical component of validity evidence" (p. xx). It is clear to us that in spite of explicit statements that competency domains and subdomains "overlap and interact with each other" (Principle 3), we must work harder to identify and elucidate this overlap with descriptions and examples.

Concluding Remarks

Our Task Force on Foundational Competencies in Educational Measurement first convened in October of 2021 and has been meeting formally and informally at every annual meeting since. There have been three symposia at NCME annual meetings, 17 formal comments on a draft, and now, eight published commentaries and a rejoinder. There are also symposia at four sibling conferences in the summer of 2024 (NCSA, the NCME Classroom Assessment Conference, ITC, and IMPS). Authors of three of these commentaries are participating in these symposia, expanding the number and breadth of perspectives. We are grateful to the authors of the eight commentaries here for their care and contributions, and we appreciate this opportunity to respond. We look forward to further dialogue in our effort to build consensus about foundational competencies in educational measurement.

References

Ackerman, T., Bandalos, D., Briggs, D., Everson, H., Ho, A. D., Lottridge, S., Madison, M., Sinharay, S.,

Rodriguez, M., Russell, M., von Davier, A., & Wind, S. (2023). *Task force response to comments*

from NCME members. National Council on Measurement in Education.

https://www.ncme.org/community/ncme-committees/educationalmeasurementtf

Ackerman, T., Bandalos, D., Briggs, D., Everson, H., Ho, A. D., Lottridge, S., Madison, M., Sinharay, S.,

Rodriguez, M., Russell, M., von Davier, A., & Wind, S. (2024). *Foundational competencies in*

educational measurement. Educational Measurement: Issues and Practice.

https://doi.org/10.3102/10769986241248771

Araneda, S., Fremer, J., & Foster, D. (2024). Commentary: Past, Present and Future of Educational

Measurement. Educational Measurement: Issues and Practice.

Borsboom, D. (2022). Possible futures for network psychometrics. *Pscyhometrika*, 87, 253-265.

https://doi.org/10.1007/s11336-022-09851-z

Briggs, D. C. (2022). NCME presidential address 2022: Turning the page to the next chapter of

educational measurement. *Journal of Educational Measurement, 59*, 398-417.

https://doi.org/10.1111/jedm.12350

Briggs, D. C. (2024). Strive for measurement, set new standards, and try not to be evil. *Journal of*

Educational and Behavioral Statistics.

https://journals.sagepub.com/doi/pdf/10.3102/10769986241238479

Brookhart, S. M. (2024). Commentary: Where does classroom assessment fit in educational

measurement? *Educational Measurement: Issues and Practice.*

Burstein, J. (2024). Responsible AI standards. Duolingo English Test. https://duolingo-

papers.s3.amazonaws.com/other/DET%2BResponsible%2BAI%2BStandards%2B-%2B040824.pdf

Castillo, W., & Gillborn, D. (2023). How to "QuantCrit": Practices and questions for education data

 researchers and users. (EdWorkingPaper: 22-546). Annenberg Institute at Brown University.

 https://doi.org/10.26300/v5kh-dd65

Cheng, Y., Pei, B., Filonczuk, A., & Le, A. T. (2024). Commentary: A data-driven analysis of recent job

 posts to evaluate the foundational competencies. *Educational Measurement: Issues and*

 Practice.

Crespo Cruz, E. J., Immanuel, A., Keller, L. A., Ketan, McIntee, K., Mena Serrano, F. J., Sireci, S. G., Smith,

 N., Suárez-Álvarez, J., Wells, C. S., Woodland, R., & Zenisky, A. L. (2024). Commentary: What is

 truly foundational? Educational Measurement: Issues and Practice.

Ebel, R. (1951). Writing the test item. In Lindquist (Ed.), *Educational Measurement* (pp. 185-249).

 American Council on Education.

Hao, J., von Davier, A. A., Yaneva, V., Lottridge, S., von Davier, M., & Harris, D. J. (2024). & Kyllonen, P.

 (2024). Transforming assessment: The impacts and implications of large language models and

 generative AI. *Educational Measurement: Issues and Practice, 43*(2), 16-29.

Ho, A. D., Ackerman, T., Bandalos, D., Briggs, D., Everson, H., Lottridge, S., Madison, M., Sinharay, S.,

 Rodriguez, M., Russell, M., von Davier, A., & Wind, S. (2023, March). *Foundational Competencies*

 in Educational Measurement. Final Report. National Council on Measurement in Education.

 https://www.ncme.org/about/news/educationalmeasurementtf

Ho, A. D. (2024). Artificial intelligence and educational measurement: Opportunities and threats. *Journal*

 of Educational and Behavioral Statistics. https://doi.org/10.3102/10769986241248771

 International Test Commission & Association of Test Publishers. (2022). Guidelines for

 technology-based assessment. https://www.intestcom.org/upload/media-library/tba-

 guidelines-final-2-23-2023-v4-167785144642TgY.pdf

Klinger, D.A., McDivitt, P.R.,Howard, B.B., Munoz, M.A., Rogers, W.T., & Wylie, E.C. (2015). *The Classroom Assessment Standards for PreK-12 Teachers*. Kindle Direct Press.

Middlebrook, K., Hamilton, L. S., & Walker, M. E. (2024). Commentary: How research and testing companies can support early-career measurement professionals. *Educational Measurement: Issues and Practice*.

National Academies of Sciences, Engineering, and Medicine. (2019). Monitoring educational equity. Washington, DC: The National Academies Press. https://doi.org/10.17226/25389

National Council on Measurement in Education. (2023). Annual Meeting Program. National Council on Measurement in Education. https://higherlogicdownload.s3.amazonaws.com/NCME/4b7590fc-3903-444d-b89d-c45b7fa3da3f/UploadedImages/NCME001-2023AnnualMeeting_Program_03302023.pdf

Raudenbush, S. W., & Sampson, R. J. (1999). Ecometrics: Toward a science of assessing ecological settings, with application to the systematic social observation of neighborhoods. *Sociological Methodology, 29*, 1–41.

Rupp, A. A. (2024). Commentary: Modernizing educational assessment training for changing job markets. *Educational Measurement: Issues and Practice*.

Schmeiser, C. B., Geisinger, K. F., Johnson-Lewis, S., Roeber, E. D., & Schafer, W. D. (1995). Code of professional responsibilities in educational measurement. National Council on Measurement in Education. https://higherlogicdownload.s3.amazonaws.com/NCME/4b7590fc-3903-444d-b89d-c45b7fa3da3f/UploadedImages/Documents/NCME_STUC__2023__Code_of_Prof_Resp_Final.pdf

Van Orman, D. S. J., Jackson, J. A., Vo, T. T., & Taylor, D. D. (2024). Commentary: Perspectives of early career professionals on enhancing cultural responsiveness in educational measurement. *Educational Measurement: Issues and Practice*.

von Davier, A. A., Mislevy, R. J., & Hao, J. (2022). Computational psychometrics: New methodologies for

a new generation of digital learning and assessment. Springer.

Wine, M., & Hoffman, A.M. (2024). Commentary: What is the breadth of "educational measurement"?

Educational Measurement: Issues and Practice.

Zumbo, B. D. (2014). What role does, and should, the test standards play outside of the United States of

America? *Educational Measurement: Issues and Practice, 33*(4), 31-33.

https://doi.org/10.1111/emip.12052