Meeting Minutes
ISBE Technical Advisory Committee for Assessment and Accountability
September 17-18, 2024 ● In-Person and Virtual

In-Person Meeting Location:
100 North First Street, Springfield, IL

Participants

ISBE	Rae Clementz, Angela Foxall, Shu-Ren Chang
TAC	Jeff Broom, Ellen Forte, Laura Hamilton, Jim Pellegrino, Mike Russell
Center	Chris Domaleski, Will Lorié, André A. Rupp, Diana Zaleski
Pearson	Mary Allen*, Tracy Gardner*, Yong Luo*
ACT	Colin Dingler**, Joanna Gorin**, Greg Hanson**, Samuel Haring, Jeremy Heneger, Mark Lewis, Joann Moore, Jay Thomas**

* Day 1 only; ** Day 2 only

Summary of TAC Recommendations

Standard-setting Work

TAC recommendations included:

- Continue efforts to recruit diverse participants to contribute to standard setting including experienced educators, content experts, students, parents, and advocates.
- Develop or refine the standard setting communication plan for the new performance standards, proactively addressing challenges to appropriate interpretation.
- Consider adjusting the current design plans for the standard-setting by rebalancing the distribution of participants to allow for tighter connections across neighboring grades for the different content areas across both days.
- Refine or change the standard-setting approach to include the best ideas from ID matching, embedded standard-setting, and modern conceptions of how qualitative information of a range of student performance is integrated into the process.
- Include a process to allow panelists in phase 2 to review the existing cut score for ACT and offer a content-based rationale for affirming or changing it.
- Consider extending the standard-setting work to include the ISA.

High School Science

TAC recommendations included:

- Revise the design of the alignment study to include PLDs, replace simpler DOK-based classifications with richer qualitative descriptions of ranges of performance using modern conceptions of cognitive complexity, and include science content experts in the process.
- Present the evolving designs for the alignment study design to the TAC so it can advise.
- Integrate cross-cutting concepts in expanded ways into future item design for ACT Science in IL.

Tuesday, September 17, 2024

ISBE Updates

Rae Clementz updated the TAC on developments since the previous meeting which included the following.

- ISBE has a fully executed contract with ACT. In connection with this, ISBE has developed a website with information about the change regarding ACT.
- Updates are being made to the Pearson platform for the IAR and ISA. These updates will not affect the student test-taking experience.
- The unified standard-setting initiative is underway. Developments since the previous TAC meeting include:
    - Policy descriptors were created and have gone out for public comment. Efforts are underway to recruit members for the policy descriptors writing team.
    - Focus groups are envisioned for outreach and will include the perspectives of stakeholders not directly involved in the writing workshop. For the panelist recruitment for summer 2025, ISBE would like to include more people with teaching experience and students.

Discussion Highlights

The TAC wondered whether other states have involved students in the standard-setting processes. The Center noted some examples but overall, student involvement is uncommon. ISBE also wants to include a stronger parent voice in this process. The TAC noted that it is important to include a diverse representation of parents (including parents not already connected to the education system through other roles, such as staff or teacher) and do targeted outreach to ensure this happens.

The TAC questioned if other states have lowered their expectations/cuts and some examples from states and programs were discussed. TAC noted that attention needs to be paid to how these issues are communicated in the general press and to be as proactive as possible. In particular, the TAC advised that it's important to keep the focus on ensuring performance expectations are well aligned with the expectations reflected in the content standards and the state's policy priorities.

ISBE noted that the superintendent has been straightforward on this issue, centering the argument that the current standards are too high. The idea is that they need to signal college and career readiness but not be so high that many students not meeting them demonstrate college and career readiness via other measures (high AP test scores, high GPAs, acceptance to colleges, etc.) thus sending conflicting messages.

The TAC suggested that ISBE consider having experienced and influential representatives attend at least some of these workshops as observers to build credibility and partner in

communication efforts. For example, some programs have invited the media to attend standard setting workshops. It may also be appropriate to invite members of the business community and the higher education space to contribute. This includes showing people example tasks and student work products that exemplify what different points on the scale represent in terms of what students know and can do. This is done effectively with NAEP items. However, some cautions were expressed that not all constituents would find this information accessible and that it is important to find multiple ways to illustrate score meaning and use.

ISBE noted that the standard-setting workshops—of which there are 10—will start by having educators review items according to difficulty and cognitive complexity. This should lead to a set of annotated items that can serve as anchors for the educators in the process. The TAC supported this work and added that it is an important opportunity to share details about the process to gain trust, which includes multiple mechanisms for communication and documentation.

Unified Standard Setting – Policy Definition (PD) Workshop

The Center and ISBE provided an update on the process and results from the Unified Standard Setting PD workshop. ISBE noted that the workshop participation was highly representative of the different stakeholders involved in this work in the state, including administrators and principals associations, two teacher's unions, regional offices of education, education policy, and advocacy groups, IL's special education advisory committee, IL's bilingual advisory committee, and student representatives. ISBE thanked Diana Zaleski and others for pulling together a really strong group of participants.

The Center reviewed the agenda of the meeting, which included a superintendent message, followed by a comprehensive presentation from ISBE. The Center provided an overview of PDs from multiple states. The bulk of the workshop was dedicated to addressing four questions that included (1) the number and names of the levels, (2) rigor, (3) policy definition, and (4) coherence. This work is summarized in a report from the workshop that was recently made available for public review.

Discussion Highlights

The Center and ISBE proposed four levels and invited the participants to react to that proposal. Most approved of having four levels as opposed to three or five. ISBE noted that the participants who advocated for three instead of four levels focused especially on the challenge of differentiating students at the low end of the scale.

The Center noted that the participants were sage about the goals of the work and not demanding too much of the state summative assessment. The TAC wondered whether this message is being communicated publicly and ISBE agreed that this is effectively captured in the current report. The TAC noted that language of performance level classifications typically gets interpreted far too deterministically, overstating what the evidence supports.

The Center highlighted a few critical points from the report such as (1) avoiding deficit language, (2) focusing on performance, not students, and (3) prioritizing clarity and utility.

The Center reviewed the conversation about rigor, with a particular note on the role of external data during the process of standard setting, including vertical articulation. This connected to a discussion on what reasonable expectations for values of SGPs would have to be (i.e., if these are too high for maintaining current proficiency status then one can question the setting of the cuts). This later led to a suggestion that these considerations could be made into a research program for validating the PDs.

The Center noted that it is very challenging to square the PD group's empirically grounded rationales with content-driven rationales. The TAC noted that the guardrails imposed by learning progressions are important, especially in certain domains such as mathematics. For example, when looking at performance at certain percentiles of the score distribution in earlier vs. later grades the differences in what students know and can do are pronounced. Moreover, learning progressions that cut across these grades are often complex and nonlinear. The TAC noted that this has design implications for the standard-setting workshop where people from different (neighboring) grades work together rather than exclusively in grade-specific groups.

The Center reiterated some insights from the discussion on coherence. Even though the participants did not settle on a singular notion of coherence, the main idea was that there are important ways in which a lack of coherence can be identified. In addition, the TAC noted that the PDs currently are not just about assessment; they are also about curriculum and instruction.

Standard Setting Plan

Pearson provided an overview of the standard setting plan for summer 2025, specifically July 14-18, 2025. Pearson underscored the collaborative nature across Pearson and ACT, which led to a joint plan. The goal of the conversation was to get feedback on the big picture mostly given that those inform certain logistical aspects that have to be settled now. In January, additional discussions will address the details of standard setting. The goal is to get feedback early to inform future planning.

Pearson noted that the plan itself is a high-level draft plan. The plan focuses on a cohesive, vertically articulated standard setting across the IAR, ISA (grades 5 and 8 moving forward), and ACT suite of assessments. The goal is to design a coherent plan that allows ACT to keep its unique value proposition for having strong validity evidence for college and career readiness. Pearson underscored that the goal is to ensure the reliability and validity of the score reporting.

Pearson proposed that all assessments, including the ACT suite, will be going through an extended, modified Angoff process for standard-setting (for Phase 1), reflecting a content-based approach. This choice is based on previous practices and seems suitable given the nature of many extended and otherwise multi-point items. It also seems to lead to a reasonable cognitive

demand for educators and links to the familiarity that they have with reviewing student work in classrooms.

Next, ACT provided an overview of the standard setting plan for the high school assessments (Phase 2). They proposed to employ a modified Briefing Book method in which panelists determine what college and career readiness imply for success in college introductory courses. In Phase 2, the primary outcome is a panelist recommendation concerning the probabilities of earning A, B, or C grades in typical, first-year credit-bearing college courses (e.g., biology for science, composition, and social science for ELA).

Discussion Highlights

Phase 1 choice of method.

- The TAC encouraged considering an alternative to extended modified Angoff, such as ID-matching; which can be incorporated via a profile-based approach that allows for reviews of extended responses for reasoning items in particular as well as the PCR tasks.
- The TAC noted that the notion of a borderline student may be difficult for participants to internalize and recommended considering an embedded standard-setting approach. Others wondered whether there would be challenges in identifying items at the lowest and highest level. This was recognized as a general challenge for ID-matching procedures but not something that prevents methods like this from being implemented with thoughtful planning.

Phase 1 design and logistics.

- The two proposed questions for panelists (one for dichotomous and another for multi-point items) are not parallel and should be reworded. This could involve a rephrasing of the second question into multiple binary decisions. Later, it was recommended to handle items with multiple points as a single set rather than distributing it across the OIB, to avoid having it re-appear multiple times.
- The TAC asked whether clusters of items were used and whether this was the intent for this process moving forward as well. Pearson confirmed that this is indeed the case as participants will review intact test forms; each form has three units with 24 MC items and 1 CR item covering a representative sample of clusters.
- The TAC wondered why the choice was made to do odd grades on Mon/Tue and even grades on Tues/Wed. Pearson noted that the goal was to include ongoing vertical articulation into this process by having panel members from Mon/Tues participate in Wed/Thurs sessions. The final vertical articulation across all grades would then be done on Friday. Another TAC member proposed another design that would accomplish the same goals while applying the current design ideas across elementary and middle school grades and, at the same time, eliminating the gap between 4th and 5th grade.

Phase 2 method rationale and design.

- The TAC viewed the ACT information as a type of impact data that is mostly a descriptive addendum to the Phase 1 decisions. It was suggested to potentially limit the

presentation of ACT data to the data at (or around) the suggested cut point from Phase 1.

- The TAC revisited the point about what specifically the outcomes of the ACT signal and to use those signals appropriately. Specifically, if students perform poorly and get a score below the cut, the messaging could be that they are not yet college ready and could benefit from additional support.

Relation between Phases 1 and 2.

- The TAC sought to clarify how the results from Phase 1 would be used in Phase 2. ACT has currently not settled on a specific approach for doing this work although they expressed openness to using the suggested cut points from Phase 1 starting points for the ACT process. Moreover, some panelists from the content-based approach will also participate in the ACT approach, representing both an opportunity for leveraging their experience and promoting coherence.
- The TAC expressed some concern that the methods seem disconnected across the two phases. Moreover, ISBE noted that the data on which the current ACT method is based is skewed to students who have self-selected into college pathways. This goes counter to the accountability signaling intentions for the high school assessment, which should go beyond college readiness.
- ACT noted that the current approach has passed federal peer review requirements in eight states and that there are additional data that will be shared with participants. ACT was asked if there was much variation in cuts used by states and responded that there was very little, mostly within one or two score points. Accordingly, it may be advisable to structure the process such that panelists are invited to affirm or challenge the proposed ACT cuts based on a content-based rationale.
- It's increasingly common to conduct a final policy review where the cut scores set by committees are reviewed by leaders from a policy perspective. This may be a useful addition to the process as a safeguard to help achieve coherence between Phases 1 and 2.

One TAC member expressed major concerns with the alignment of the ACT to the state standards / NGSS, which would impact not just the logic underlying the standard-setting intention but also the practical implementation of the standard-setting process. As a back-up, consider implementing a standard setting workshop for the ISA.

Wednesday, September 18, 8:30-11:30 CDT

ACT Science

ACT provided an overview of the science test with a focus on evidence of alignment to the IL science standards. In addition, procedures for item development, form construction, scoring, reporting, and accessibility were discussed. The team addressed specifically how it made

different design decisions for the assessment based on NGSS and research at ACT. They noted that their assessment includes exclusively MC items and does not include technology-enhanced items. An alignment study is included as part of the current contract.

Discussion Highlights

The TAC asked which states are currently using the ACT Science assessment. States include: Alabama, Montana, Nebraska, and Wisconsin. There have been no adaptations to the specific state contexts.

A TAC member clarified that the states currently using ACT Science have not adopted NGSS. Among the states that have adopted the NGSS, only a few use ACT Science but none use the assessment as the state science assessment for ESSA. TAC members expressed concern that ACT science could meet the requirements of peer review for NGSS states. In particular, evidence of alignment will be challenging. A submission of peer review evidence is being planned for January 2026, both for the ISA and ACT Science in IL.

The TAC strongly recommended an improved design for the alignment study. The TAC noted that alignment studies for science are sometimes done relatively poorly. They reiterated that the blueprint should specifically match the state expectations and is a representative sampling. As a result, PLDs should be part of the alignment study. Moreover, the TAC recommended that DOK should not be used as part of that process. Instead, the specific standards-based expectations for processing requirements, both in terms of the stimuli and the response format, should be used in a more detailed manner. These decisions intimately connect to the score reporting and interpretation work.

A TAC member pointed to the work by New Meridian as an example. Moreover, alignment should elicit qualitative feedback about the rationale for decisions and criteria for sufficiency.

The TAC also advocated for the inclusion of science content experts during the alignment workshops as well as having more detailed information about the proposed alignment work. The TAC could partner with ISBE in this work as desired.

The TAC also asked about the degree to which grade-specific content knowledge is required to solve the problems. Generally speaking, the contexts appear possibly too generic based on the sample items shown. ACT noted that the design does require the knowledge representative of common courses in 9th grade for a certain percentage of items even though separate subscores are not reported at this level. ACT already has - and is continuing to expand - a body of evidence that demonstrates the relevance of content knowledge for the contexts presented in the items.

The TAC discussed whether the adoption of ACT should signal a shift in course offerings and expectations. The TAC underscored the careful considerations that had been given to the connection between variation in course offerings, content knowledge expectations, and

associated signaling functions in IL when the ISA was designed. ACT noted that national research has shown very different course patterns for different kinds of students in the high school space. This leads to differential performance for students with different pathways on the same items, which sometimes creates psychometric challenges.

The connection between instruction and assessment is a critical issue, which the teacher's association has spoken to in detail in advance of the meeting. The ACT Science assessment seems incongruent with the current instructional practices.

The TAC wondered about the relationship of performance on items that required detailed content knowledge with information about the contexts in which they have been exposed to this knowledge. ACT noted that they recently did a series of cognitive labs with about 30 students, which showed that 10th graders did better than seniors on many aspects even though they were otherwise excelling. They also did an extensive analysis of all content-based items from the last 10 years, which showed that many items failed more traditional statistical tests due to low discrimination indices. This suggests that unidimensional scaling approaches are challenging for these items.

Cross-cutting concepts are indeed an essential part of NGSS and need to be foregrounded in educational practices. Hence, it is important to integrate them into the item design.

The TAC has previously discussed student choice in assessment which could be further explored as a mechanism to allow for personalization around the DCI. This could be done using a two-stage setup conceptually, although this is very challenging to implement in practice given that both online and paper forms with comparable properties are desired. ACT noted that the constraints are indeed the major roadblocks that currently lead to some of the complicated discussions. ACT underscored that it is advisable to consider designing a system of assessments with summative and classroom components.

The TAC asked about the degree to which the ACT Science assessment can meet the accommodation requirements for IL comprehensively; in particular, the concern was whether there are stipulations in IL law that would invalidate an ACT score. ISBE noted that neither vendor was open to transadaptation of the assessment. There are plans to embed TTS functionalities that would preclude the use of TTS if used outside of formal accommodation requirements.

ISBE noted that a major concern from the teacher's perspective is the difference between the current ISA and the ACT Science assessment expectations around vocabulary, especially the higher cognitive load for the ACT Science. However, ISBE also qualified that this concern may, in fact, be more about the instructional realities on the ground and the associated increases in expectations. In response, ACT noted that all of their assessments are developed with a universal design framework.

It is important to collect information about instructional practices and opportunity to learn. ISBE emphasized that their instruction and curriculum team for science is engaged in targeted improvement efforts that are already ongoing. It is unclear what kinds of data are specifically collected about these efforts at scale. ISBE noted that there is a challenge with variability in course taking and participation in the science assessment, which was one of the drivers of choosing the ACT Science assessment.

TAC Planning Session

This session began with a round-robin discussion among all participants. A majority of the reflections were about standard setting. Highlights are provided without specific attributions to individuals.

Standard setting

Some of the standard-setting processes could be streamlined, especially beyond the distribution of people and especially regarding the cognitive framing of the tasks. In connection with this, it is important to leverage the core expertise of the panelists, which is in classroom experience. This specifically implies moving away from the probabilistic judgments arising from various Angoff-type methods.

Concerning the distribution of teachers across grades, teachers are capable of thinking across grade boundaries generally and there is indeed sufficient variability across IL to sample for this kind of setup. The TAC suggested leveraging/sharing the resources that are created as part of this process for future professional development.

Consider what kinds of judgments panelists can make in Phase 2 while avoiding conceptual and practical incoherence with Phase 1.

Consider how PLDs for science are generated so that they are coherent given that assessments are only being used in select grades while standards exist for most grades.

Adoption of ACT

Consider the importance of thinking through the consequences of implementing the ACT Science assessment for instructional and professional development practices.

Alignment Studies

Expectations for the alignment study for ACT Science are high and go beyond the IL contexts as a successful implementation will likely be used by ACT to approach other potential clients whose assessments need to be aligned to the NGSS explicitly.

Consider sufficient initial buy-in on the design of the study (in much the same way that some research studies are pre-registered) to make sure there is sufficient initial buy-in and agreement about the design and approach.

Alignment with Instructional Practices

Consider a more detailed tracking of the breadth, depth, and quality of implementation of curriculum, instruction, and professional development efforts across the state to guide the development of reasonable expectations for performance changes on state assessments.

The next meeting was confirmed to be on January 22-23, 2025.

Closed Session - IAR Review

Summary of TAC Recommendations

TAC recommendations included:

- Consider producing SGP evidence supporting the score increase in a composite graphic Including additional data breakdowns by race/ethnicity jointly with socioeconomic status (and possibly gender), considering available sample sizes in the resulting groups.
- Including descriptions of the performance of former ELLs and, additionally, look at the reading and writing performance on the ELP assessments.

IAR Review (closed session to ISBE, TAC, and Pearson)

Pearson provided an overview of preliminary, uncorrected assessment results from 2024, reviewing patterns of performance by grade, content area, and student group. Pearson also discussed efforts to ensure a representative sample of students encounter embedded field test (EFT) items considering the influence of text-to-speech availability on form-taking patterns. Pearson then provided an update on item and form development for 2025.

Pearson noted that it wanted to give a detailed overview of the results to allow all TAC members and ISBE to have access to all relevant information without a specific agenda to discuss particular issues in this meeting.

Discussion Highlights

In terms of the overall scores, there have been several positive gains across content areas and grades, which ISBE noted is also reflected in the Student Growth Percentiles (SGPs). In connection with this, the TAC suggested displaying the SGP evidence supporting the score increase in a composite graphic, which would be more suitable for technical audiences only. Among other benefits, this information would provide additional support for the credibility of outcomes to help assess converging evidence that the score increases are real.

Overall, the increases are less pronounced in math, especially in the upper grades. The TAC noted that this makes sense given that math is more cumulative in terms of knowledge acquisition than ELA so earlier losses or deficiencies are more challenging to recover in later grades.

The science performance showed relatively steady performance across the years. However, the TAC noted that they questioned whether these outcomes are fully trustworthy given that the standards were originally developed under PARCC but the ISA is a different kind of assessment, which confounds interpretation of results. There was discussion among TAC members about the discrepancy in percent proficient between science and the other subjects,

with suggestions that they are out of alignment. When considering IL students' performance on NAEP science, the result is potential confusion among report audiences. The TAC noted that the unified standard-setting process that is currently being planned in conjunction with a new science assessment should address these issues.

Results by student groups

- Females showed higher subgroup performance than males on the scale score and percent proficient metric for ELA. For math, the pattern is not as uniform, and differences are less pronounced. It was noted that these patterns are reflective of national trends, especially when it comes to writing, which makes up 50% of the blueprint.
- Group differences by socioeconomic status for both ELA and math are strong and mirror those from NAEP. The TAC noted that it would be valuable to run additional breakdowns by race jointly with socioeconomic status (and possibly gender). The same pattern is observed for disability status in both content areas.
- Group differences for ELLs in ELA are also pronounced, with many ELL students showing very poor performance. ISBE noted that they are indeed tracking former ELLs and but that newcomers to the country are tracked separately. ISBE could also differentiate between learners who have been in the system for different years. Pearson noted the extensive amount of writing that they need to do based on synthesizing evidence from multiple sources. The TAC suggested that ISBE could consider separately reporting the performance of former and 'ever' ELLs to further include descriptions of the performance of former ELLs. Moreover, analyzing performance separately for the reading and writing domains may provide additional insights about group performance.
- With respect to group differences by race/ethnicity, the TAC noted that intersections should be looked at as there have been other cases of meaningful differences (e.g., Asian students from lower SES are often performing particularly poorly.) While sample sizes have to be watched in additional breakdowns, the TAC underscored that the SES and race/ethnicity intersectionality is particularly important but that further intersectionalities with gender should be considered if possible.

The TAC wondered how the data are used during reporting and decision-making to help better understand intended and actual interpretations. ISBE noted that there is a high-level deck that is shared with information from the core report card. Districts are then allowed to conduct additional analyses, and many districts indeed conduct additional analyses. ISBE releases a state-level de-identified student-level file with some reasonable appropriate suppressions to support additional analysis for small sample sizes and no geographic information.

The Center noted that it would be helpful to visualize conditional improvement (e.g., by deciles), which would illuminate performance patterns for students at difference places in the distribution (e.g., are improvements or declines occurring primarily for high or low performing students or are patterns uniform?).

The next section of the presentation focused on text-to-speech (TTS) accommodations. The key issue is how TTS items could be handled for equating; they are currently excluded.

There are 30+ forms but only one online form has the TTS accommodation. This form was taken by 30% of students whereas 70% of students took the other 30 forms, which contained field test items. As a result, these forms contained smaller sample sizes than ideal, and the students who took the TTS form possibly had unique demographic characteristics that were not represented in the equating process. This reduced sample size for about 30 forms causes issues with items that have multiple score points, as some score points will have scarce or missing data. Next year, the plan is to have ten forms with TTS accommodations and, in the long run, all forms are planned to have a TTS variant.

Pearson noted that all TTS forms have the same operational items, but using more TTS forms allows for broader field testing. Science will also become all TTS. ISBE noted that the assignment of TTS is currently done at the building level, but in the future, with every form having TTS, students could also self-select TTS accommodations. ISBE confirmed that there will be Spanish-language TTS versions for math.

ISBE has provided guidance that suggests that this accommodation should only be provided to students whenever it is part of their daily instructional practice. They have heard that this is now a common practice, and so the accommodation is given relatively commonly. This means that it is turned on for these students but not necessarily selected by them. 30% of students who were offered TTS use it.

There was a discussion about the risk of giving too much autonomy to students, given that not all may be equally informed about their selection. Some research shows that selection can differently impact performance for higher versus lower performing students. The Center noted that research supports that these kinds of selection affordances are best for students on the upper end of achievement rather than students at the lower end who might get unduly distracted.

Pearson described a sensitivity analysis of including the data from the TTS forms in the post-equating process using the operational items; they were calibrated and then converted back to the baseline parameters using the Stocking-Lord procedure. Graphs with individual item parameters showed that their estimates are very similar across estimation contexts.

The Center noted that the ICCs are jointly determined by the item parameters given the use of the two-parameter model. The Center also wondered whether there are any plans to review items once they have been identified. Pearson noted that the ICC analysis is valuable but it would likely be challenging still to identify problematic items unless effects are small. The team also underscored that the most important perspective is to look at scale score differences and performance levels at the aggregate.

The TAC discussed the importance of reviewing both item stability and overall scale score differences when evaluating equivalence. It's important not to dismiss differences in the upper and lower regions of the scale, further from the cutscores, because precision throughout the scale is important to support growth.

The Center cautioned against dismissing (larger) differences for individual items. The TAC also supported this perspective given that one could imagine certain item types that might have design features that create systematic differences - for example vocabulary items with homonyms or number items.

Pearson noted that there are very small differences in scale scores, with a few one- or two-point differences at points of the scale that are relatively far away from cut scores. It was noted that the differences in the raw-to-scale conversations between accommodated and non-accommodated forms were comparable to those typically found between parallel forms. The TAC generally supported these interpretations and noted that this could be evidence in support of federal peer review (element 5).

Pearson noted that review cycles have internally led to improvements of the tagging processes for items, especially to support better quality-control monitoring for automated test assembly processes. The TAC wondered whether this could lead to removal of a class of items, thus possibly leading to systematic construct underrepresentation. In response, Pearson elaborated that construct coverage is closely monitored and feedback is used to improve item quality and promote fairness.

ISBE noted that these ideas are similar to accommodations for alternate forms. The TAC also noted a connection between this work and Braille forms. They also noted that experiences of students with low-incidence disabilities also experience classroom environments very differently.

ISBE noted that there are possibilities to be more purposeful in the content and bias review processes during early (and later) reviews, including cross-pollinating the respective working groups through joint membership. The TAC suggested having some of these colleagues write specifications as well.

The next section of the discussion was focused on automated test assembly. Pearson noted that the best approach involved maximizing test information at the proficiency cut and minimizing test information differences across forms while maintaining the same raw cut score across the four cuts.

The TAC noted that the current analyses are centered around the proficiency cut that currently exists rather than the future cuts that will be created through the unified standard-setting process.

Pearson reminded participants that only one form will be used for the standard-setting process next year while impact data will be computed across all operational forms. They underscored

that there are complex rules in place that guide the automated item assembly that were used to derive the current results.

It's important to ensure the forms used for standard setting are sufficiently 'granular.' The Center underscored cautions around granularity, noting that gaps in the form selection can affect standard-setting outcomes even if the forms have overall identical raw score cut points and very similar ICCs. Depending on the method chosen, certain methods such as the Bookmark method can interact with the gaps on the scale between ordered items. One solution is to construct an augmented form for standard setting.

A discussion ensued about the possibility of using artificial intelligence methods to generate proposed items with specifications. Pearson noted that internal research is ongoing.