Assessing the Options

Considerations for provision of choice in assessment

smarter BALANCED

BEYOND STANDARD

Assessing the options:

Considerations for provision of choice in assessment

Submitted to the Smarter Balanced Assessment Consortium, by
Paulina Biernacki, Stanford University

March 10, 2022

Acknowledgements: The author would like to extend special thanks to Magda Chia, Richard Durán, James Pellegrino, Maria Araceli Ruiz-Primo, and Guillermo Solano-Flores for their thought partnership and guidance. The author would also like to thank Audrey Lesondak and Vitaliy Shyyan, who provided very thoughtful feedback on earlier drafts of this report. Any errors, however, are solely attributed to the author of this brief.

Suggested citation:

Biernacki, P. J. (2022). *Assessing the options: Considerations for provision of choice in assessment*. Smarter Balanced Assessment Consortium.

https://portal.smarterbalanced.org/library/en/practice-brief-assessing-the-options.pdf

Executive Summary

Allowing students to make choices (e.g., on the reading materials to read for an assignment) can contribute to creating learning environments that promote critical thinking, and intrinsic motivation among students. Given this potential that choice has to enhance instruction, it is reasonable to examine its potential as a contributor to enhanced assessment practices.

Choice in assessment—allowing students to make decisions on the ways in which they are assessed within a given array of tasks (including stimulus materials), response formats, and grading schemes or scoring systems—has the potential to mitigate the fact that, in addition to the knowledge being assessed, multiple individual factors, including experience, learning styles, anxiety, stigma, and disability shape student performance on assessments.

Fairness is an important motivation for providing choice in assessment, as it can potentially remove barriers that may prevent students from demonstrating knowledge. To achieve this ideal, it is important for education agencies and educators to have realistic expectations about the challenges, possibilities, and limitations of providing choice in assessment at each of three levels of a balanced assessment system—classroom assessment, school-district assessment, and state assessment for accountability.

Assessments at the classroom level (both formative and summative) are closer to the students' personal experiences. Therefore, using choice in classroom assessment (and especially formative assessment activities), can potentially provide teachers with more accurate information on their students' learning. In contrast, in large-scale assessment, choice may pose a validity threat to the extent that it compromises comparability since the condition of standardization—the condition in which all test takers are tested under the same conditions – would be compromised.

Provision of choice in assessment at any level is based on the premise that students are well equipped with a set of critical thinking and metacognitive skills that are relevant to effective choice-making. In large-scale assessment, provision of choice in addition poses multiple logistical and challenges that are difficult to surmount with currently available psychometric theory and methods.

This brief is organized into five sections:

1. **Introduction**
2. **Types of choice in assessment**
3. **Choice in assessment: Goals and challenges**
4. **Conclusions**
5. **Recommendations for education agencies and educators**

1 Introduction

In an information era characterized by rapid shifts in technology and learning environments, students need to be supported in learning how to ask questions, seek information, navigate new tools for learning, think critically, and become self-directed learners.[1, 2] As a pedagogical resource that promotes autonomy,[3] choice appears to be critical to meeting such goals.

Research from the learning sciences indicates that, in supportive learning contexts, provision of choice can strengthen conceptual understanding, memory, and application of knowledge.[4, 2] Research also shows that choice can promote student autonomy and intrinsic motivation—the desire to engage in a given behavior regardless of any external rewards.[1] Autonomy and motivation are core aspects of individual self-directed growth that is associated with persistence, engagement, and task performance.[5, 6] For example, there is evidence that students who are allowed to choose between two options in their homework tasks can be more motivated to complete homework, describe themselves as more competent on the homework, and demonstrate higher achievement on tests than students who are not allowed to choose between homework tasks.[6] Also, there is evidence that students who are allowed to choose among several reading materials can be more engaged, report greater control over their reading, and appear to demonstrate deeper learning than their peers who are not allowed to choose.[7]

There is evidence that students who are allowed to choose between two options in their homework tasks can be more motivated to complete homework.

However, in and of itself, the provision of choice is not a guarantee of effective instruction. There is evidence that offering too many choices can actually reduce engagement.[5] For example, a meta-analysis that compiled evidence from various studies on the effects of choice on task performance in children and adults found that two to four choices appeared to be optimal, where individuals could select from an array of three to five options per choice.[8] This principle may vary depending on the learning contexts, the content being taught, the characteristics of the tasks, and the characteristics of the students.

Given the potential to support learning, it is reasonable to ask how choice can contribute to producing accurate information about individual student learning in assessment. *Choice* in assessment is the term used to refer to the condition in which students are allowed to select the ways in which they are assessed within a given array of types of tasks and their stimulus materials (e.g., a variety of reading passages), response formats (e.g., multiple-choice vs open-ended items), or grading schemes or scoring systems.

Examining the potential of providing choice in assessment offers the possibility to devise ways of supporting students to best demonstrate their knowledge in assessments by allowing them to select optimal testing according to their personal experiences, learning styles, and preferences. This adaptability can potentially contribute to the development of balanced assessment systems[9] by optimizing the information on student achievement and minimizing measurement error attributable to student factors that are irrelevant to the knowledge assessed. The challenges needed to meet this ideal are different for each level of assessment (classroom, school or district, state).[10, 11] Summative assessment (assessment *of* learning) takes place at the classroom, school-district, and state levels. In classroom assessment (and especially formative assessment or assessment *for* learning), the provision of choice can potentially render accurate information on student learning that teachers can use to enrich their instruction. In contrast, in large-scale assessment, the provision of choice may pose validity threats and compromises standardization—the condition in which all test takers are tested under the same conditions, and therefore, comparability.

The goal of this brief is to provide state and district education leaders and educators with some considerations about the possibilities and challenges in providing choice in assessment. Since educational systems are vast and complex, the recommendations in this brief are not offered as a panacea for eliminating achievement gaps across students. These gaps are regarded as stemming from a complex web of issues that include teacher preparation, school resources, and policy challenges that are beyond what any form of assessment can do.

2 Types of choice in assessment

Choice may involve different kinds of decisions students need to make. A simple way of examining the different forms of choice offered to students can be examined according to the components that define an assessment—task, response format, and scoring system.[12] Table 1 provides examples of potential choices that could be done using the assessment components and categories provided in other sources. This is not an exhaustive list of types of choice in assessment; also assessment choices could be grouped into sets of categories different from the three assessment components.[13, 14, 15] The table is intended to give an idea of the variety of choices that can be provided in assessment and to provide a simple framework to reason about the ways in which choice can involve different aspects of assessments.

Evidence supporting the use of choice in classroom assessment has focused on tasks (e.g., choices of reading passages or writing prompts). Only budding research on choice of response format and scoring exists. While this research has demonstrated some promising increases in student engagement and satisfaction in learning,[15, 16, 17] research is needed that addresses the cases in which choice may undermine student achievement for students who do not understand how to make advantageous decisions or when too many choices are offered.

Table 1: Examples of Potential Choices by Assessment Component.

Assessment Component	Examples
Task: *An observation assumed to be a sample that is representative of a knowledge domain that will provide the evidence to draw reasonable inferences about what students know.*	*The student selects the type of task/item or stimulus materials that tap the targeted domain-specific knowledge constructs to work with.* • **Item format.** Select, generate, explain, produce, demonstrate (e.g., select predictions, generate predictions, explain predictions or select examples of concepts, generate examples of concepts, explain why examples reflect concept attributes). • **Stimulus materials.** Context in which the content of the item is embedded (e.g., the reading passage or the topic in an English Language Arts test). • **Item sample.** Select from a targeted construct a set of items considered parallel and equivalent (e.g., four items to respond to, out of a menu of six items).
Response Format: *Medium used to capture student's performance.*	*The student selects how to provide their responses.* • **Mode.** Computer-based, paper-and-pencil, orally. • **Response format aligned to the item format.** Select from options, write an essay, fill-in the blank, etc.
Grading Scheme/ Scoring System: *Method for making sense of the evidence provided by the student's performance.*	*The student selects from a set of rules for systematically assigning a grade to their response based on the relevance of the evidence to an interpretation.* • **Elimination of tasks/items.** Eliminate one item from the test or one assignment from the final grade. • **Inclusion of tasks/items.** Adding an item or completing a bonus project to adjust grading.

3 Choice in assessment:
Goals and challenges

Fairness is an important motivation for examining the possibilities of choice in assessment, as it has the potential to remove barriers that may prevent students from demonstrating knowledge. In addition to knowledge of the content being assessed, individual backgrounds and local contexts influence how students interpret test items and how they respond to them.[18] Additional factors such as anxiety or stigma may impact performance on tests and prevent students from demonstrating their knowledge in full.[14]

While standardization—a basic condition for comparability—is intended to ensure the same testing conditions for all test takers,[19] it does not control for student background differences and the disparate sets of learning opportunities schools made available to students. Ideally, assessment tasks can be developed in ways that are sensitive to large numbers of students' characteristics. This process of development is based on *understanding* student group heterogeneity and individual student needs, discerning how student characteristics interact with test characteristics, and seeking ways to redesign tests so that the testing process can more flexibly meet students' needs.[20] It can be argued that, whereas sameness in standardized testing can promote *equality*, flexibly promote *equity* in testing. According to this reasoning, students may be better equipped to demonstrate their knowledge with tasks that assess the same sets of constructs in different ways.

The implications of choice in classroom assessment mainly concern the benefit (e.g., engagement and depth of learning) of allowing students to make choices, for example, among a set of tasks (e.g., essay versus oral presentation) or between grading schemes.[16, 17] In contrast, the implications of choice in large-scale assessment are logistical and technical. First, offering choice increases assessment development work and costs, as it entails creating more tasks, response formats, and scoring systems for the same given test. It also involves the development of complex analytical approaches that allow proper treatment of data to develop a common measurement scale across the multiple forms of a test that result from the different combinations of tasks or response formats.[21]

While standardization—a basic condition for comparability—is intended to ensure the same testing conditions for all test takers,[19] it does not control for student background differences and the disparate sets of learning opportunities schools made available to students.

Second, the provision of choice in assessment may compromise score comparability across students.[19, 21] Whereas standardized assessments require fixed testing conditions and tasks, choice introduces flexibility and differences in task difficulty that cannot be fully controlled and accounted for. There is evidence that student performance is unstable across types of tasks[22]. Also, the interaction of student and type of task (e.g., multiple-choice items versus constructed written response) is a major source of measurement error[23]. The validity of score interpretations may be compromised also if tasks of different types intended to assess the same construct are wrongly assumed to be equivalent.[14]

Third, the validity of score interpretations may be compromised also if it is wrongly assumed that all test takers are able to make the decisions that lead them to best demonstrate their knowledge or that they have given the opportunities to develop the metacognitive skills needed to self-monitor as they navigate choice-making decisions. Students are not necessarily clear about what they know and what they do not know; there is evidence that they cannot predict accurately their performance in tests.[24] The National Assessment of Educational Progress (NAEP) reader study conducted by Campbell and Donahue,[25] which compared the performance of students who were and were not allowed to select reading passages in reading tasks, found that such choices did not improve students' scores, and, in certain cases, actually led to inequitable score differences based on gender and ethnicity.[21, 25] Moreover, students may lack the confidence and self-efficacy needed to make decisions, especially if they are stigmatized and alienated in school (e.g., if placed in remedial programs).[26]

The diminishing return of choice may relate to burdens on cognitive demands. Too many choices may increase cognitive demands, placing undue strains on working memory. Based on knowledge from the field of cognitive science,[27] it is possible to anticipate that the cognitive load of an item may increase unnecessarily if the examinee needs to use their working memory to make many choices that are not directly related to completing the task at hand.

Too many choices may increase cognitive demands, placing undue strains on working memory.

4 Conclusions

Within the landscape of *twenty-first century skills*, choice can potentially contribute to promoting higher-order thinking skills (depending on the choice selected), including metacognition, and helping students to become more autonomous learners.[2, 28] Available evidence indicates that, these skills are critical to successfully make choices in assessment. Thus, provision of choice in assessment may be not fair or effective if students are underprepared to make advantageous decisions or if they are only allowed to make choices when they are assessed. Therefore, students need guidance for making choices on assessment items and tasks that are paralleled in everyday learning. Classroom is the best environment to promote educated choices. It is critical that these choices are situated within a supportive culture of autonomy and choice-making in the classroom, where such choices are aligned to recognized learning goals and objectives.

At the classroom level, teachers are best positioned to personalize assessments in ways that align to students' individual needs, curricular materials, and local contexts. For assessment at the school or district level and at the state level, the situation is less clear. Given the current theory and methods in educational measurement, the potential advantages of choice do not appear to outweigh the disadvantages that result from compromising comparability. At the school or district level and the state level, it is difficult to account for individual student needs and personal contexts. Students' choices are difficult to implement in large-scale testing and can lead to difficulties in score interpretation,[19, 21] where it is not always possible to control for important variables that impact choice-making, such as metacognitive skills, and self-efficacy.

Research is needed to examine how assessment systems can be redesigned to foreground student diversity, account for student-level factors, and address the interaction of those student-level factors with testing. While the provision of choice raises issues of score comparability, it has been suggested that "understandardized" assessment tasks can more fully capture student knowledge and skills. As Sireci[20] argues, "the small sacrifice in test score comparability may lead to great gains in test score validity" (p.5). Research is needed that examines how testing practices need to be revised so that assessment tasks can more flexibly meet students' needs.

In sum, while in principle, choices involved in tasks, response formats, and scoring systems can be integrated at the classroom, district, and state levels, in practice, currently, it seems to be most feasible to provide choices at the classroom level. Still, even in classroom-based assessment, the implicit assumption that students have the proper metacognitive skills that allow them to make good choices may not always hold. At the very least, the provision of assessment choice at any of the three assessment levels makes it necessary that educators provide supportive guidance that allows students to make wise choices. Provision of choice in assessment is not a panacea and may backfire if not used cautiously.

5 Recommendations
for education agencies and educators

1 State education agencies should support research that examines the conditions that best support the development of metacognition and self-regulation among students and their promotion in the classroom, as critical to choice-making in both learning and assessment. State education agencies also should support research that examines ways in which choice can be used coherently within assessment systems in ways that both support learning.

2 State education agencies and local education agencies should partner with professional development (PD) organizations to help teachers promote a culture of autonomy in their classrooms wherein students' choices converge upon common, recognized learning goals and targets. These efforts should support teachers in appreciating both the value of promoting student autonomy in learning and the need for effective strategies for implementation that are flexible and adaptable to different student age groups, content areas, local contexts, and student needs. Ideally, classroom cultures of autonomy should allow students to have a sense of comfort and familiarity with making choices in their everyday classroom activities. Such comfort and familiarity in everyday learning tasks would then facilitate choice making processes in assessment tasks.

3 State education agencies and local education agencies should partner to help teachers explicitly support the development of self-regulation and metacognition. Explicit instruction can help students make appraisals, increase self-efficacy, and strengthen critical thinking skills to minimize sources of measurement error in choice-making. In tandem with instruction on other general test-taking strategies, teachers can empower students to more fully express their knowledge, skills, and abilities when they are assessed. This is not a recommendation for general test preparation regarding test content or an encouragement to teach test taking skills at the cost of subject instructional time, but rather for helping students understand testing and develop appropriate choice-making strategies.

4 State education agencies should evaluate their messaging to local education agencies regarding the ideas of choice and assessment. Specifically, close attention should be paid to the messaging regarding the relationship between choice and equity and the similarities and differences in choice across classroom-, district-, and state-levels of assessment.

5 State leaders should partner with assessment experts to ensure that innovations introduced in assessment systems and assessment procedures concerning choice are sufficiently supported by empirical evidence on their effectiveness. Consideration should be given not only to the impact of choice on students' test scores, but also to the impact on teacher use of assessment information, as well as student affect and higher-order thinking skills (e.g., engagement, motivation, metacognition). It is vital that this work be carried out with direct participation of students and educators.

6 State leaders should partner with assessment experts to systematically examine the effectiveness of choice on various student subgroups. Some students may be adversely impacted by provision of choice if they are less familiar with making choices. Such adverse effects may be especially pronounced among those who are at risk of poverty or stigma and among students from cultures that do not encourage autonomy among young children.

References and Notes

[1] Wineburg, S., McGrew, S., Breakstone, J., & Ortega, T. (2016). *Evaluating Information: The Cornerstone of Civic Online Reasoning*. Stanford Digital Repository. http://purl.stanford.edu/fv751yt5934

[2] Darling-Hammond, L., Flook, L., Cook-Harvey, C., Barron, B. & Osher, D. (2020) Implications for educational practice of the science of learning and development, *Applied Developmental Science, 24*(2), 97-140, https://doi.org/10.1080/10888691.2018.1537791

[3] Haworth, L. (1986). *Autonomy: an essay in philosophical psychology and ethics*. New Haven: Yale University Press.

[4] Pretorius, L., van Mourik, G. P., & Barratt, C. (2017). Student choice and higher-order thinking: Using a novel flexible assessment regime combined with critical thinking activities to encourage the development of higher order thinking. *International Journal of Teaching and Learning in Higher Education, 29*(2), 389–401

[5] Iyengar, S.S., & Lepper, M.R. (2000). When a choice is demotivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology, 79*(6), 995-1006. http://doi.org/10.1037//0022-351479.6.995

[6] Patall, E. A., Cooper, H., & Wynn, S. R. (2010). The effectiveness and relative importance of choice in the classroom. *Journal of Educational Psychology, 102*(4), 896-915. http://dx.doi.org.stanford.idm.oclc.org/10.1037/a0019545

[7] Morgan, D. N., & Wagner, C. W. (2013). "What's the catch?": Providing reading choice in a high school classroom. *Journal of Adolescent & Adult Literacy, 56*(8), 659–667. http://www.jstor.org/stable/41827920

[8] Patall, E. A., Cooper, H., & Robinson, J. C. (2008). The effects of choice on intrinsic motivation and related outcomes: A meta-analysis of research findings. *Psychological Bulletin, 134*(2), 270–300. https://doi.org/10.1037/0033-2909.134.2.270

[9] Marion, S., Thompson, J., Evans, C., Martineau, J., Dadey, N., & National Center for the Improvement of Educational Assessment, I. (NCIEA). (2019). *The challenges and opportunities of balanced systems of assessment: A policy brief*. National Center for the Improvement of Educational Assessment. https://www.nciea.org/library/challenges-and-opportunities-balanced-systems-assessment-policy-brief

[10] Herman , J. L. , Wilson , M. R. , Shavelson , R. , Timms , M. , & Schneider , S. (2005, April). *The CAESL assessment model*. Paper presented at American Educational Research Association annual conference, Montreal, Canada.

[11] Pellegrino, J. W., Chudowsky, N., Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.

[12] Ruiz-Primo, M. A. & Shavelson, R. J. (1996). Rhetoric and reality in science performance assessments: An update. *Journal of Research in Science Teaching, 33*(10), 1045-1063.

[13] Bennett, R. E. (1993). On the meanings on constructed response. In. R. E. Bennett & W. C. Ward (Eds.), Construction versus choice in cognitive measurement: Issues in constructed response, performance testing, and portfolio assessment (pp. 1-27). Hillsdale, NJ, 1993.

[14] Snow, R. E. (1993). Construct validity and constructed-response tests. In. R. E. Bennett & W. C. Ward (Eds.), Construction versus choice in cognitive measurement: Issues in constructed response, performance testing, and portfolio assessment (pp. 45-60). Hillsdale, NJ, 1993.

[15] Rose, D. H., Robinson, K. H., Hall, T. E., Coyne, P., Jackson, R. M., Stahl, W. M., & Wilcauskas, S. L. (2018). Accurate and informative for all: Universal design for learning (UDL) and the future of assessment. In S. N. Elliott, R. J. Kettler, P. A. Beddow, & A Kurz, (Eds.). (2018). *Handbook of accessible instruction and testing practices: Issues, innovations and applications*. New York: Springer

[16] Garside, J., Nhemachena, J. Z., Williams, J., & Topping, A. (2009). Repositioning assessment: Giving students the 'choice' of assessment methods. *Nurse Education in Practice, 9*(2), 141–148. https://doi.org/10.1016/j.nepr.2008.09.003

[17] Gosselin, J., & Gagné, A. (2014). *Differentiated evaluation: An inclusive evaluation strategy aimed at promoting student engagement and student learning in undergraduate classrooms*. Toronto, Canada: Higher Education Quality Council of Ontario. https://heqco.ca/wp-content/uploads/2020/03/Differentiated_Evaluation_ENG.pdf

[18] Solano-Flores, G., & Li, M. (2009). Generalizability of cognitive interview-based measures across cultural groups. *Educational Measurement Issues and Practices, 28*(2), 9-19.

[19] Berman, A. I., Haertel, E. H., & Pellegrino, J. W. (2020). *Comparability of large-scale educational assessments: Issues and recommendations*. Washington, DC: National Academy of Education.

[20] Sireci, S.G. (2020). Standardization and UNDERSTANDardization in educational assessment. *Educational Measurement: Issues and Practice, 39*(3) 100-105. https://doi.org/10.1111/emip.12377

[21] Wainer, H., & Thissen, D. (1994). On examinee choice in educational testing. *Review of Educational Research, 64*(1), 159–195. https://doi.org/10.3102/00346543064001159

[22] Ruiz-Primo, M. A., Baxter, G. P., Shavelson, R. J. (1993). On the stability of performance assessments. *Journal of Educational Measurement, 30*(1), 41-53.

[23] Shavelson, R. J., Baxter, G. P., & Pine, J. (1992). Performance assessments: Political rhetoric and measurement reality. *Educational Researcher,* 21(4), 22-27.

[24] Winnie, P. H., & Azevedo, R. (2014). Metacognition. In R. K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (2nd ed., pp. 63-87). New York, NY: Cambridge University Press.

[25] Campbell, J. R., & Donahue, P. L. (1997). *Students selecting stories: The effects of choice in reading assessment: Results from The NAEP Reader special study of the 1994 National Assessment of Educational Progress*. U.S. Dept. of Education, Office of Educational Research and Improvement, Educational Resources Information Center.

[26] Flammer, A. (1995). Developmental analysis of control beliefs. In A. Bandura (Ed.) *Self-Efficacy in Changing Societies*. (pp. 69-115). Cambridge: Cambridge University Press.

[27] Sweller, J. (2005). Implications of cognitive load theory for multimedia learning. In R. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (Cambridge Handbooks in Psychology, pp. 19-30). Cambridge: Cambridge University Press.

[28] Pellegrino, J. W. & Hilton, M. L. (Eds.). (2012). *Education for life and work: Developing transferable knowledge and skills in the 21st century*. National Academies Press.