

Appendix Validation Article Review Instrument

Introduction

Based on the *Standards for Educational and Psychological Testing*, developed by the American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME) and hereafter referred to as the *Standards* (AERA et al., 2014), the work of Kane (2013) and Messick (1995), and other scholars in the field of validity, this instrument is designed to support article reviews with an eye toward issues of validity in general, and specifically toward the validation of particular interpretations and uses of SAT and ACT test scores.

The sections and items in this instrument follow the framework shown in Figure 1. Simply stated, each of the wedges represent different sources of evidence that either support or challenge the specific interpretation/use argument (IUA; see Kane, 2013) being validated. The IUA relies on the quality of the scores that inform it; thus, the five sources of evidence pictured support the IUA through the measurement and methodology issues on which the IUA rests.

Figure 1

Framework for Validity Evidence



Citation: Amrein-Beardsley, A., Azizova, Z. T., Gibbs, N. P., Ikegwuonu, E., Kim, J., La Torre, D. M., Lavery, M. R., Pivovarova, M., & Zheng, Y. (2025). A validation review of the SAT and ACT for college and university admissions decisions. *Education Policy Analysis Archives*, 33(28).
<https://doi.org/10.14507/epaa.33.8734>

Reviewer

Which member of the research team is completing this entry?

- Reviewer 1 (Blinded for peer review)
- Reviewer 2 (Blinded for peer review)
- Reviewer 3 (Blinded for peer review)
- Reviewer 4 (Blinded for peer review)
- Reviewer 5 (Blinded for peer review)
- Reviewer 6 (Blinded for peer review)
- Reviewer 7 (Blinded for peer review)
- Reviewer 8 (Blinded for peer review)

First Author

Copy/paste or type the last, first name of the sole or first author here.

Potential Conflict of Interest (COI)

Is any contributing author affiliated with the SAT/College Board or the ACT?

- Yes (1)
- No (2)
- Unsure (3)

Year of Publication

Copy/paste or type the four-digit year of the publication here.

Full Title

Copy/paste or type the full title of the publication here.

Citation: Amrein-Beardsley, A., Azizova, Z. T., Gibbs, N. P., Ikegwuonu, E., Kim, J., La Torre, D. M., Lavery, M. R., Pivovarova, M., & Zheng, Y. (2025). A validation review of the SAT and ACT for college and university admissions decisions. *Education Policy Analysis Archives*, 33(28).
<https://doi.org/10.14507/epaa.33.8734>

Type of Publication

What type of publication is the reviewed piece?

- Report of Empirical Research (1)
- Systematic Review or Meta-Analysis (2)
- Program Evaluation Report (3)
- Methodological, Theoretical, or Conceptual Piece (4)
- Other (type publication classification below) (5)

Purpose of the Study

What is the original purpose of the study reviewed (which may or may not be directly related to the IUA)?

This should be 1) copied/pasted from the abstract, 2) copied/pasted from another part of the article, or 3) written in simple terms by the reviewer if not stated briefly by the author(s).

Methodological Approach

What general methods were used in the study reviewed?

This should be 1) copied/pasted from the abstract, 2) copied/pasted from another part of the article, or 3) written in simple terms by the reviewer if not stated briefly by the author(s).

General Findings

What are the findings reported in the paper (which may or may not pertain to the IUA under investigation, but should pertain to the purpose listed in 0.6)?

This should be 1) copied/pasted from the abstract, 2) copied/pasted from another part of the article, or 3) written in simple terms by the reviewer if not stated briefly by the author(s).

Citation: Amrein-Beardsley, A., Azizova, Z. T., Gibbs, N. P., Ikegwuonu, E., Kim, J., La Torre, D. M., Lavery, M. R., Pivovarova, M., & Zheng, Y. (2025). A validation review of the SAT and ACT for college and university admissions decisions. *Education Policy Analysis Archives*, 33(28).
<https://doi.org/10.14507/epaa.33.8734>

1.0 Interpretation/Use Argument (IUA)

“To validate an interpretation or use of test scores is to evaluate the plausibility of the claims based on the test scores. Validation therefore requires a clear statement of the claims inherent in the proposed interpretations and uses of the test scores. Public claims require public justification” (Kane, 2013, p.1). The Interpretation/Use Argument (IUA) "includes all of the claims based on the test scores (i.e., the network of inferences and assumptions inherent in the proposed interpretation and use)" (Kane, 2013, p. 2). "The kinds of evidence required for validation are determined by the claims being made, and more-ambitious claims require more evidence than less-ambitious claims" (Kane, 2013, p. 3). "More-ambitious interpretations (e.g., a construct interpretation or a causal claim) tend to be more useful than less-ambitious interpretations, but they are harder to validate" (Kane, 2013, p. 37).

Via this review we will address how the reviewed article contributes to the following IUA:

SAT and ACT tests accurately, consistently, and fairly measure college readiness and can predict college success; thus, SAT and ACT scores support valid college admissions decisions.

- College readiness is defined as having mastered the prerequisite knowledge and skills to participate in traditional college coursework without remediation.
- College success is defined as persistence (i.e., students return the semester following their first semester), retention (i.e., students return the fall following their first year), program completion (i.e., students complete all requirements for their degree program), and degree attainment (e.g., graduation).
- The SAT and ACT scores that inform the IUA are defined as those that are used (or to be used) to predict a student’s college readiness and aptitude for college success, collectively and particularly in the following subject areas: English/reading and mathematics, as well as science, writing, history, and foreign languages, which are often optional.
- Valid college admissions decisions are defined as follows: For first-time, freshmen, seeking entrance into the University of California (UC) System, as well as other selective institutions, indicators pursued for admissions emphasize goals pertaining to:
 - Selectivity (e.g., keeping admission and matriculation rates lower/higher than previous years, as weighted at least in part by lower/higher high school grade point averages (GPAs), SAT/ACT scores, and the like); and
 - Equity & Access (e.g., enrollment percentages of students by first-generation status, race/ethnicity, Pell-grant reciprocity, out-of-state or international status). If UC (and other selective institutions) try to balance and appropriately weigh admission indicators around these two goals, for purposes of this study we define a “valid decision” to be based on one’s

Citation: Amrein-Beardsley, A., Azizova, Z. T., Gibbs, N. P., Ikegwuonu, E., Kim, J., La Torre, D. M., Lavery, M. R., Pivovarova, M., & Zheng, Y. (2025). A validation review of the SAT and ACT for college and university admissions decisions. *Education Policy Analysis Archives*, 33(28).
<https://doi.org/10.14507/epaa.33.8734>

demonstrated “academic excellence” as balanced by applicants’ relevant background factors such as those listed prior.

- Note that the IUA concerns the validity of individual admissions and/or scholarship decisions; decisions that are made at the level of the applicant. Selectivity and diversity, as concerns at the institutional level, are therefore most likely to be represented in the Validity Evidence Based on Related Consequences area of this instrument. Equity and access are terms most often used to reflect the applicant-level side of the diversity coin. For example, when an individual student from a traditionally underrepresented group is inappropriately or unfairly denied admission, it is described as an equity and access problem but influences the institution’s diversity (or lack thereof). Similarly, college readiness and college success (defined above) are terms that describe the applicant-level of the selectivity coin. When students who are college ready and likely to be successful in college are inappropriately or unfairly denied admission, it may preserve the institution’s selectivity at the expense of the individual qualified to attend. Because the commonly accepted definition of selectivity includes consideration of SAT/ACT scores, selectivity must only be considered in the Validity Evidence Based on Related Consequences area of this instrument.
- More specifically, as per UC Regents Policy 2102 (see [here](#)):
 - The undergraduate admissions policy of the UC system is guided by the UC system’s commitment to serve the people of California and the needs of the state.
 - The entrance requirements established by the UC system requires that the top one- eighth of the state's high school graduates, as well as those transfer students who have successfully completed specified college work, be eligible for admission to the UC system. These requirements are designed to ensure that all eligible students are adequately prepared for university-level work.
 - Mindful of its mission as a public institution, the UC system has an historic commitment to provide places within each UC campus for all eligible applicants who are residents of California. The UC system seeks to enroll, on each of its campuses, a student body that, beyond meeting eligibility requirements, demonstrates high academic achievement or exceptional personal talent, and that encompasses the broad diversity of cultural, racial, geographic, and socioeconomic backgrounds characteristic of California.
 - Because applicant pools differ among the campuses of the UC system, each campus shall establish procedures for the selection of applicants to be admitted from its pool of eligible candidates. Such procedures shall be consistent with the principles stated above and with other applicable UC policies.
 - Note that it may be helpful to complete the items in this section last.

1.1 Position on IUA being Validated

What kind of position do the author(s) of the publication reviewed take on the IUA above?

Note: This is the author(s)' take on the IUA as revealed through the publication reviewed, not the reviewer's assessment of how well the publication supports the IUA in question.

Citation: Amrein-Beardsley, A., Azizova, Z. T., Gibbs, N. P., Ikegwuonu, E., Kim, J., La Torre, D. M., Lavery, M. R., Pivovarova, M., & Zheng, Y. (2025). A validation review of the SAT and ACT for college and university admissions decisions. *Education Policy Analysis Archives*, 33(28).
<https://doi.org/10.14507/epaa.33.8734>

- (5) SUPPORTIVE (The publication proposes or supports the IUA) (1)
- (4) SOMEWHAT SUPPORTIVE (The publication is generally supportive of the IUA, though it may advise some caution) (2)
- (3) NEITHER SUPPORTIVE NOR CHALLENGING (The publication provides evidence or discussion in support of the IUA as well as evidence or discussion which challenges it) (3)
- (2) SOMEWHAT CHALLENGING (The publication critiques, challenges, or offers caution about the IUA while acknowledging its appropriate use) (4)
- (1) CHALLENGING (The publication challenges, opposes, or argues against the IUA) (5)
- (88) PERIPHERAL (The investigated IUA is not a central aspect of the publication, but the publication contains evidence relevant to the review) (6)
- (99) NOT APPLICABLE (The article reviewed provides no relevant evidence related to the IUA being investigated. Specifically, none of the numbered areas below are addressed) (7)

1.2 IUA-Related Findings

What (if any) findings are reported in the paper that directly pertain to the IUA under investigation?

This should be 1) copied/pasted from the abstract, 2) copied/pasted from another part of the article, or 3) written in simple terms by the reviewer if not stated briefly by the author(s).

1.3 IUA-Related Recommendations

What (if any) specific recommendations do the authors make in the paper that directly pertain to the IUA under investigation?

These may be recommendations to use SAT/ACT scores in admissions decisions, recommendations to cease doing so, or cautions to consider. This should be 1) copied/pasted from the abstract, 2) copied/pasted from another part of the article, or 3) written in simple terms by the reviewer if not stated briefly by the author(s).

Citation: Amrein-Beardsley, A., Azizova, Z. T., Gibbs, N. P., Ikegwuonu, E., Kim, J., La Torre, D. M., Lavery, M. R., Pivovarova, M., & Zheng, Y. (2025). A validation review of the SAT and ACT for college and university admissions decisions. *Education Policy Analysis Archives*, 33(28).
<https://doi.org/10.14507/epaa.33.8734>

1.X IUA Notes

Use this space for notes and comments about how the publication reviewed addresses the IUA under investigation that are not captured by any other section in this review instrument.

Remember to include quotes and page numbers for any text copied and pasted from the article reviewed.

2.0 Reliability/Precision and Fairness

2.1.0 Reliability/Precision and Stability

The authors of the *Standards* (AERA et al., 2014) point out that the measurement literature has historically used the term “reliability/precision to denote the more general notion of consistency of the scores across instances of the testing procedure” (p. 33). Though discussed “as an independent characteristic of test scores, . . . reliability/precision has implications for validity” (AERA et al., 2014, p. 34), also as disaggregated, whereby “to the extent feasible (i.e., if samples sizes are large enough), reliability/precisions should be estimated separately for all relevant subgroups (e.g., defined in terms of race/ethnicity, gender, language proficiency) in the population” (AERA et al., 2014, p. 37; see also Standard 2.3).

Specifically, evidence that SAT/ACT scores are stable (or unstable) over repeated administrations address this area of the framework. Cronbach’s alpha values are measures of internal consistency and are only relevant here if reported for the entire test as an indicator of reliability. Cronbach’s alpha values reported for subtests (such as mathematics, reading, science, etc.) are more relevant to area 5, validity evidence based on internal structure.

2.1.1 Addressed

To what extent is reliability/precision or stability addressed within the publication reviewed?

- (3) DIRECTLY STUDIED (evidence gathered/reported as part of the publication) (1)
- (2) MENTIONED WITH EVIDENCE (e.g., cites other papers) (2)

Citation: Amrein-Beardsley, A., Azizova, Z. T., Gibbs, N. P., Ikegwuonu, E., Kim, J., La Torre, D. M., Lavery, M. R., Pivovarova, M., & Zheng, Y. (2025). A validation review of the SAT and ACT for college and university admissions decisions. *Education Policy Analysis Archives*, 33(28).
<https://doi.org/10.14507/epaa.33.8734>

- (1) MENTIONED (no citations or evidence provided; mark "NA" below and skip to the next numbered area of evidence) (3)
- (0) NOT ADDRESSED (mark "NA" below and skip to the next numbered area of evidence) (4)

2.1.2 Evidence

If DIRECTLY STUDIED or MENTIONED WITH EVIDENCE, does the publication provide evidence that supports or challenges the IUA under investigation?

- (5) Evidence SUPPORTS the IUA under investigation (1)
- (4) Evidence SOMEWHAT SUPPORTS the IUA under investigation (2)
- (3) Evidence NEITHER SUPPORTS NOR CHALLENGES the IUA under investigation (3)
- (2) Evidence SOMEWHAT CHALLENGES the IUA under investigation (4)
- (1) Evidence CHALLENGES the IUA under investigation (5)
- (99) (NA) Not Applicable (0 or 1 selected above) (6)

2.1.X Notes

Use this space for notes and comments related to reliability/precision or stability. Remember to include quotes and page numbers for any text copied and pasted from the article reviewed.

2.2.0 Fairness

The *Standards* (AERA et al., 2014) discusses fairness as “a fundamental issue in protecting test takers and test users in all aspects of testing” and that “fairness is a fundamental validity issue and requires attention throughout all stages of test development and use” (p. 49). In its glossary, the *Standards* define fairness as follows: The validity of test score interpretations for intended use(s) for individuals from all relevant subgroups. A test that is fair minimizes construct- irrelevant variance associated with individual characteristics and testing contexts that would otherwise compromise the validity of scores for some individuals. (AERA et al., 2014, p.219)

Construct irrelevant variance (CIV) is a term used by Messick (1989) to describe factors that falsely inflate or deflate the measurement of a variable and therefore distort its interpretation, or distort its

Citation: Amrein-Beardsley, A., Azizova, Z. T., Gibbs, N. P., Ikegwuonu, E., Kim, J., La Torre, D. M., Lavery, M. R., Pivovarova, M., & Zheng, Y. (2025). A validation review of the SAT and ACT for college and university admissions decisions. *Education Policy Analysis Archives*, 33(28).
<https://doi.org/10.14507/epaa.33.8734>

validity (see also AERA et al., 2014, p. 12). CIV is pertinent here in that the presence of CIV “systematically lowers or raises scores for identifiable groups of test takers and results in inappropriate score interpretations,” as often based on students’ opportunities to learn that “can influence the fair and valid interpretations of test scores for their intended users” (AERA et al., 2014, p. 54). Authors of the *Standards* (AERA et al., 2014) discuss several aspects of fairness, many of which will be captured in other areas of this instrument. *Standards* authors discuss measurement bias as a major threat for fairness, along with issues of accessibility and universal design.

Note: The operational distinction that we make for the purpose of this study is that evidence has been gathered as part of a reviewed study for the explicit purpose of determining whether the test performs similarly for various subgroups of test takers, some of whom might be considered vulnerable, disadvantaged, or traditionally underrepresented in colleges and universities (e.g., first-generation college students, students with disabilities, students from low SES backgrounds, students for whom the English language is not their first language or their home language) would be captured in section 2.2, Fairness. In these cases, the author(s) will typically use words like “fairness,” “equity,” “access,” “discrimination,” or “social justice” in their narrative. If evidence that could be captured in another area of this instrument is collected and analyzed without reference to such issues, then it should be recorded in that portion of the instrument.

2.2.1 Addressed

To what extent is fairness addressed within the publication reviewed?

- (3) DIRECTLY STUDIED (evidence gathered/reported as part of the publication) (1)
- (2) MENTIONED WITH EVIDENCE (e.g., cites other papers) (2)
- (1) MENTIONED (no citations or evidence provided; mark "NA" below and skip to the next numbered area of evidence) (3)
- (0) NOT ADDRESSED (mark "NA" below and skip to the next numbered area of evidence) (4)

2.2.2 Evidence

If DIRECTLY STUDIED or MENTIONED WITH EVIDENCE, does the publication provide evidence that supports or challenges the IUA under investigation?

- (5) Evidence SUPPORTS the IUA under investigation (1)
- (4) Evidence SOMEWHAT SUPPORTS the IUA under investigation (2)
- (3) Evidence NEITHER SUPPORTS NOR CHALLENGES the IUA under investigation (3)
- (2) Evidence SOMEWHAT CHALLENGES the IUA under investigation (4)

Citation: Amrein-Beardsley, A., Azizova, Z. T., Gibbs, N. P., Ikegwuonu, E., Kim, J., La Torre, D. M., Lavery, M. R., Pivovarova, M., & Zheng, Y. (2025). A validation review of the SAT and ACT for college and university admissions decisions. *Education Policy Analysis Archives*, 33(28).
<https://doi.org/10.14507/epaa.33.8734>

- (1) Evidence CHALLENGES the IUA under investigation (5)
- (99) (NA) Not Applicable (0 or 1 selected above) (6)

2.2.X Notes

Use this space for notes and comments related to fairness. Remember to include quotes and page numbers for any text copied and pasted from the article reviewed.

3.0 Validity Evidence Based on Test Content

Evidence based on test content is primarily concerned with the relationship between the content of the tests and the constructs the tests are purported to measure (Cronbach, 1971; Cureton, 1951; Kane, 2013). “Content-oriented evidence of validation is at the heart of the process in the educational arena known as alignment, which involves evaluating the correspondence between student learning standards and test content” (AERA et al., 2014, p. 15). Accordingly, validity evidence of this type may include but not be limited to analyses of the content domain as aligned to the tested domain, discussions of the relevance or sufficiency of the tested content domains as pertinent to the proposed interpretations, or recommendations and testimonials taken from content-area experts (AERA et al., 2014; Kane, 2013).

Note: As a validation review, evidence may only be included in this section of the instrument if the author(s) of the paper being reviewed addresses the alignment of the SAT or ACT test scores to effectively measure college readiness and college aptitude as defined prior; that is, the IUA of interest herein. Furthermore, this area of evidence will only be "directly studied" or "mentioned with evidence" if it relies on more than just logical argument but presents empirical or external evidence to support the argument.

3.1 Addressed

To what extent is content-related evidence of validity addressed within the publication reviewed?

- (3) DIRECTLY STUDIED (evidence gathered/reported as part of the publication) (1)
- (2) MENTIONED WITH EVIDENCE (e.g., cites other papers) (2)
- (1) MENTIONED (no citations or evidence provided; mark "NA" below and skip to the next numbered area of evidence) (3)

Citation: Amrein-Beardsley, A., Azizova, Z. T., Gibbs, N. P., Ikegwuonu, E., Kim, J., La Torre, D. M., Lavery, M. R., Pivovarova, M., & Zheng, Y. (2025). A validation review of the SAT and ACT for college and university admissions decisions. *Education Policy Analysis Archives*, 33(28).
<https://doi.org/10.14507/epaa.33.8734>

- (0) NOT ADDRESSED (mark "NA" below and skip to the next numbered area of evidence) (4)

3.2 Evidence

If DIRECTLY STUDIED or MENTIONED WITH EVIDENCE, does the publication provide evidence that supports or challenges the IUA under investigation?

- (5) Evidence SUPPORTS the IUA under investigation (1)
- (4) Evidence SOMEWHAT SUPPORTS the IUA under investigation (2)
- (3) Evidence NEITHER SUPPORTS NOR CHALLENGES the IUA under investigation (3)
- (2) Evidence SOMEWHAT CHALLENGES the IUA under investigation (4)
- (1) Evidence CHALLENGES the IUA under investigation (5)
- (99) (NA) Not Applicable (0 or 1 selected above) (6)

3.X Notes

Use this space for notes and comments related to content related evidence of validity. Remember to include quotes and page numbers for any text copied and pasted from the article reviewed.

4.0 Validity Evidence Based on Response Processes

The Response Processes source of evidence captures how well the construct or performance of the target domain matches the responses or performances measured in order to inform the IUA (AERA et al., 2014). When the responses or performances measured by a test are substantially different from that of the target domain, irrelevant method variance may limit the validity of proposed interpretations of the test (Messick, 1989, 1995). Evidence in this area may include but not be limited to investigations into how test takers interpret and respond to the test used, logical or empirical analyses of the match between response processes and the target domain, or studies of raters and graders and the processes that they employ (AERA et al., 2014). In the Response Processes area, we considered Kane's (2013) suggestion that, "if the [measured] tasks seem to involve the same processes as most tasks in the target domain, extrapolation is likely to seem reasonable" (p. 28). Specifically, if the proposed interpretations and uses of SAT and ACT scores are college admissions decisions, then the knowledge, skills, and processes that are necessary to succeed

Citation: Amrein-Beardsley, A., Azizova, Z. T., Gibbs, N. P., Ikegwuonu, E., Kim, J., La Torre, D. M., Lavery, M. R., Pivovarova, M., & Zheng, Y. (2025). A validation review of the SAT and ACT for college and university admissions decisions. *Education Policy Analysis Archives*, 33(28).
<https://doi.org/10.14507/epaa.33.8734>

in college should be the same knowledge skills and processes that are necessary to succeed on these tests.

Note: Evidence related to the “coachability” of the SAT and ACT might belong in this source of validity evidence. If SAT/ACT preparation were only to involve reviewing and mastering the content measured by the test, then it would be relevant to area 3 above, validity evidence based on test content. If SAT/ACT preparation activities involve anything other than academic content (e.g., test-taking strategies, time management during the test), then it likely belongs here.

4.1 Addressed

To what extent is response process related evidence addressed within the publication reviewed?

- (3) DIRECTLY STUDIED (evidence gathered/reported as part of the publication) (1)
- (2) MENTIONED WITH EVIDENCE (e.g., cites other papers) (2)
- (1) MENTIONED (no citations or evidence provided; mark "NA" below and skip to the next numbered area of evidence) (3)
- (0) NOT ADDRESSED (mark "NA" below and skip to the next numbered area of evidence) (4)

4.2 Evidence

If DIRECTLY STUDIED *or* MENTIONED WITH EVIDENCE, does the publication provide evidence that supports or challenges the IUA under investigation?

- (5) Evidence SUPPORTS the IUA under investigation (1)
- (4) Evidence SOMEWHAT SUPPORTS the IUA under investigation (2)
- (3) Evidence NEITHER SUPPORTS NOR CHALLENGES the IUA under investigation (3)
- (2) Evidence SOMEWHAT CHALLENGES the IUA under investigation (4)
- (1) Evidence CHALLENGES the IUA under investigation (5)
- (99) (NA) Not Applicable (0 or 1 selected above) (6)

4.X Notes

Use this space for notes and comments related to response processes. Remember to include quotes and page numbers for any text copied and pasted from the article reviewed.

Citation: Amrein-Beardsley, A., Azizova, Z. T., Gibbs, N. P., Ikegwuonu, E., Kim, J., La Torre, D. M., Lavery, M. R., Pivovarova, M., & Zheng, Y. (2025). A validation review of the SAT and ACT for college and university admissions decisions. *Education Policy Analysis Archives*, 33(28).
<https://doi.org/10.14507/epaa.33.8734>

5.0 Validity Evidence Based on Internal Structure

In the historical validity literature, evidence based on internal structure is related to construct validation (see Cronbach & Meehl, 1955) and concerns the degree to which the test that supports the IUA reflects the theory on which it is based. This includes evidence of how faithfully the sub-scales, sub-tests, or various components of the testing procedure follow their hypothesized relationships. “Analyses of the internal structure of a test can indicate the degree to which the relationships among test items and test components conform to the construct on which the proposed test score interpretations are based” (AERA et al., 2014, p. 16).

Typically, evidence based on internal structure is psychometric in nature and concerns the degree to which the test that supports the IUA reflects the theoretical construct on which it is based. This can include evidence of how faithfully the factor analyses, sub-scales, sub-tests, or various components of the testing procedure follow their hypothesized relationships. Given that the SAT and ACT both produce subtest scores as well as overall scores, this area of validity evidence might include examinations of those subtests and/or their relationships to the other subtests.

5.1 Addressed

To what extent is internal structure addressed within the publication reviewed?

- (3) DIRECTLY STUDIED (evidence gathered/reported as part of the publication) (1)
- (2) MENTIONED WITH EVIDENCE (e.g., cites other papers) (2)
- (1) MENTIONED (no citations or evidence provided; mark "NA" below and skip to the next numbered area of evidence) (3)
- (0) NOT ADDRESSED (mark "NA" below and skip to the next numbered area of evidence) (4)

5.2 Evidence

If DIRECTLY STUDIED or MENTIONED WITH EVIDENCE, does the publication provide evidence that supports or challenges the IUA under investigation?

- (5) Evidence SUPPORTS the IUA under investigation (1)
- (4) Evidence SOMEWHAT SUPPORTS the IUA under investigation (2)

Citation: Amrein-Beardsley, A., Azizova, Z. T., Gibbs, N. P., Ikegwuonu, E., Kim, J., La Torre, D. M., Lavery, M. R., Pivovarova, M., & Zheng, Y. (2025). A validation review of the SAT and ACT for college and university admissions decisions. *Education Policy Analysis Archives*, 33(28).
<https://doi.org/10.14507/epaa.33.8734>

- (3) Evidence NEITHER SUPPORTS NOR CHALLENGES the IUA under investigation (3)
- (2) Evidence SOMEWHAT CHALLENGES the IUA under investigation (4)
- (1) Evidence CHALLENGES the IUA under investigation (5)
- (99) (NA) Not Applicable (0 or 1 selected above) (6)

5.X Notes

Use this space for notes and comments related to response processes. Remember to include quotes and page numbers for any text copied and pasted from the article reviewed.

6.0 Validity Evidence Based on Relations to Other Variables (e.g., convergent, concurrent, discriminant, or predictive)

In many cases, the intended interpretation for a given use implies that the construct should be related to some other variables, and, as a result, analyses of the relationship of test scores to variables external to the test provide another important source of validity evidence” (AERA et al., 2014, p. 16). Relationships between the measurement of interest and other variables (e.g., performance criteria) have been historically discussed as a primary aspect of criterion validity (see Cronbach, 1971; Kane, 2013; Moss, 1992, 1995). Such evidence may can be experimental or correlational; concurrent or convergent, demonstrating the degree to which two measures of constructs that theoretically should be related are, in fact, related; discriminant, demonstrating the degree to which concepts or measurements that are supposed to be unrelated are, in fact, unrelated; or predictive, concerning the degree to which measurement output can be used to predict other outcomes, which are typically assessed or observed at some later point in time. In all cases, researchers typically seek evidence related to other variables to demonstrate that the IUA is consistent with conclusions based on other related measures, observations, or outcomes. The extent to which such evidence can be generalized to new situations or multiple types of test takers is in large measure a function of the accumulated research.

6.1 Addressed

To what extent are relationships with other variables addressed within the publication reviewed?

- (3) DIRECTLY STUDIED (evidence gathered/reported as part of the publication) (1)
- (2) MENTIONED WITH EVIDENCE (e.g., cites other papers) (2)

Citation: Amrein-Beardsley, A., Azizova, Z. T., Gibbs, N. P., Ikegwuonu, E., Kim, J., La Torre, D. M., Lavery, M. R., Pivovarova, M., & Zheng, Y. (2025). A validation review of the SAT and ACT for college and university admissions decisions. *Education Policy Analysis Archives*, 33(28).
<https://doi.org/10.14507/epaa.33.8734>

- (1) MENTIONED (no citations or evidence provided; mark "NA" below and skip to the next numbered area of evidence) (3)
- (0) NOT ADDRESSED (mark "NA" below and skip to the next numbered area of evidence) (4)

6.2 Evidence

If DIRECTLY STUDIED or MENTIONED WITH EVIDENCE, does the publication provide evidence that supports or challenges the IUA under investigation?

- (5) Evidence SUPPORTS the IUA under investigation (1)
- (4) Evidence SOMEWHAT SUPPORTS the IUA under investigation (2)
- (3) Evidence NEITHER SUPPORTS NOR CHALLENGES the IUA under investigation (3)
- (2) Evidence SOMEWHAT CHALLENGES the IUA under investigation (4)
- (1) Evidence CHALLENGES the IUA under investigation (5)
- (99) (NA) Not Applicable (0 or 1 selected above) (6)

6.3 Types

If DIRECTLY STUDIED or MENTIONED WITH EVIDENCE, what types of evidence based on relationships to other variables did the publication reviewed provide (select all that apply)?

- CONCURRENT/CONVERGENT EVIDENCE - The degree to which measures of the same or similar constructs are related (1)
- DISCRIMINANT EVIDENCE - The degree to which measures of different constructs are unrelated (2)
- PREDICTIVE EVIDENCE - The degree to which a test or observation accurately predicts future performance or future outcomes (these future criteria might not be tests per se, but might be outcomes such as persistence, retention, four-year graduation, etc.) (3)

6.X Notes

Use this space for notes and comments related to relationships with other variables. For each piece of evidence based on relations to other variables, be sure to include (a) the variable for which a relationship with SAT/ACT scores was tested, (b) the reported coefficient (or range of coefficients), (c) the significance or non-significance of the relationship, and (d) the effect size as reported/interpreted by the author(s). Remember to include quotes and page numbers for any text copied and pasted from the article reviewed.

Citation: Amrein-Beardsley, A., Azizova, Z. T., Gibbs, N. P., Ikegwuonu, E., Kim, J., La Torre, D. M., Lavery, M. R., Pivovarova, M., & Zheng, Y. (2025). A validation review of the SAT and ACT for college and university admissions decisions. *Education Policy Analysis Archives*, 33(28).
<https://doi.org/10.14507/epaa.33.8734>



7.0 Validity Evidence Based on Related Consequences

If the inferences and decisions that the scores support must be validated, rather than the scores themselves, then it is appropriate to also evaluate the intended and unintended consequences of those score uses as part of the validity argument. Whether the IUA produces its intended outcomes is an important consideration, as is whether the IUA yields unintended outcomes (that can be positive or negative) as well as, or in lieu of, those which are intended. As per AERA et al. (2014) Standard 1.25: “When unintended consequences result from test use, an attempt should [also] be made to investigate whether such consequences arise from the test’s sensitivity to characteristics other than those it is intended to assess or from the test’s failure to fully represent the intended construct. (p. 30). After all, Kane (2013) writes that “a decision rule that achieves its goals at an acceptable cost and with acceptable consequences is considered a success. A decision rule that does not achieve its goals or has unacceptable consequences is considered a failure” (p. 47). Note, however, that not all consequences will be immediately apparent and that, particularly in the case of unintended consequences, a particular IUA might be in use for some time before the full extent of its consequences are known and understood. Subsequently, validity evidence based on related consequences is the element of the framework most likely to reopen a scholarly discussion that might have previously been seen as settled or resolved. Appropriately, it is often observed that “the validation process never ends” (AERA et al., 2014, p. 21).

To help reviewers calculate the balance of intended and unintended consequences reported by the literature reviewed, the three categories provided by Kane (2013) will be split into four separate notes fields: 7.A) INTENDED OUTCOMES, 7.B) POSITIVE SYSTEMIC EFFECTS (unintended), 7.C) NEGATIVE SYSTEMIC EFFECTS (unintended), 7.D) DIFFERENTIAL EFFECTS / ADVERSE IMPACT (unintended)

7.1 Addressed

To what extent are intended and/or unintended consequences of the IUA under investigation addressed within the publication reviewed?

- (3) DIRECTLY STUDIED (evidence gathered/reported as part of the publication) (1)
- (2) MENTIONED WITH EVIDENCE (e.g., cites other papers) (2)
- (1) MENTIONED (no citations or evidence provided; mark "NA" below and skip to the next numbered area of evidence) (3)
- (0) NOT ADDRESSED (mark "NA" below and skip to the next numbered area of evidence) (4)

Citation: Amrein-Beardsley, A., Azizova, Z. T., Gibbs, N. P., Ikegwuonu, E., Kim, J., La Torre, D. M., Lavery, M. R., Pivovarova, M., & Zheng, Y. (2025). A validation review of the SAT and ACT for college and university admissions decisions. *Education Policy Analysis Archives*, 33(28).
<https://doi.org/10.14507/epaa.33.8734>

7.2 Evidence

If DIRECTLY STUDIED or MENTIONED WITH EVIDENCE, does the publication provide evidence that supports or challenges the IUA under investigation?

- (5) Evidence SUPPORTS the IUA under investigation (1)
- (4) Evidence SOMEWHAT SUPPORTS the IUA under investigation (2)
- (3) Evidence NEITHER SUPPORTS NOR CHALLENGES the IUA under investigation (3)
- (2) Evidence SOMEWHAT CHALLENGES the IUA under investigation (4)
- (1) Evidence CHALLENGES the IUA under investigation (5)
- (99) (NA) Not Applicable (0 or 1 selected above) (6)

7.A INTENDED OUTCOMES

Use this space for notes and comments related to the intended consequences of the IUA under investigation. Remember to include quotes and page numbers for any text copied and pasted from the article reviewed.

7.B POSITIVE UNINTENDED EFFECTS (applies to system(s) or population(s))

Use this space for notes and comments related to positive unintended consequences of the IUA under investigation that affect everyone. Remember to include quotes and page numbers for any text copied and pasted from the article reviewed.

7.C NEGATIVE UNINTENDED EFFECTS (applies to system(s) or population(s))

Use this space for notes and comments related to negative unintended consequences of the IUA under investigation that affect everyone. Remember to include quotes and page numbers for any text copied and pasted from the article reviewed.

Citation: Amrein-Beardsley, A., Azizova, Z. T., Gibbs, N. P., Ikegwuonu, E., Kim, J., La Torre, D. M., Lavery, M. R., Pivovarova, M., & Zheng, Y. (2025). A validation review of the SAT and ACT for college and university admissions decisions. *Education Policy Analysis Archives*, 33(28).
<https://doi.org/10.14507/epaa.33.8734>

7.D DIFFERENTIAL EFFECTS / ADVERSE IMPACT (applies to individuals or individuals by subpopulation(s))

Use this space for notes and comments related to unintended consequences of the IUA under investigation that affect certain sub-groups or populations differently than others. Remember to include quotes and page numbers for any text copied and pasted from the article reviewed.

8.0 Other Validity Notes

“A sound validity argument integrates various strands of evidence into a coherent account of the degree to which existing evidence and theory support the intended interpretation of test scores for specific uses” (AERA et al., 2014, p. 21).

8.1 Overall Rating

According to the **reviewer**, and based upon the readings and discussions of the research team, taken in its entirety, how strong is the evidence of validity provided by this publication for the IUA being investigated in this review?

- (5) SUPPORTIVE (1)
- (4) SOMEWHAT SUPPORTIVE (2)
- (3) NEITHER SUPPORTIVE NOR CHALLENGING (3)
- (2) SOMEWHAT CHALLENGING (4)
- (1) CHALLENGING (5)

8.A Author's Concerns

What concerns or limitations do the author(s) raise about the IUA that have not already been captured in another part of this form?

Citation: Amrein-Beardsley, A., Azizova, Z. T., Gibbs, N. P., Ikegwuonu, E., Kim, J., La Torre, D. M., Lavery, M. R., Pivovarova, M., & Zheng, Y. (2025). A validation review of the SAT and ACT for college and university admissions decisions. *Education Policy Analysis Archives*, 33(28).
<https://doi.org/10.14507/epaa.33.8734>

8.B Reviewer's Concerns

What concerns or limitations does the reviewer see that have not already been captured by another part of this form?

8.X Other Notes

Provide other thoughts, comments, and notes here about the publication in its entirety or about issues not addressed elsewhere in this form.

0.X Star Rating

On a five-star scale, how would you rate the overall value of this publication for inclusion on a list of recommended reading for those interested in the use of SAT and ACT tests in college admissions decisions?

- (5) Five Stars (1)
- (4) Four Stars (2)
- (3) Three Stars (3)
- (2) Two Stars (4)
- (1) One Star (5)