

## 21<sup>st</sup> Century Skills Assessment Scoring System FAQs

### 1. What is the significance of the 4 levels of proficiency?

**Answer:**

**Below Basic:** The student is lacking basic skills for their grade level.

**Basic:** The student has the skills that are appropriate for entering their grade level, but is not prepared for moving to the next grade level (6<sup>th</sup> grade or 9<sup>th</sup> grade).

**Proficient:** The student is ready to advance to the next grade level.

**Advanced:** The student has skills that are beyond the next grade level.

### 2. How was the proficiency standard set?

**Answer:** The proficiency benchmarks for students were created using the 24 ISTE NETS-S 2007 performance indicators and NETS-S profiles. When standards serve as educational goals, they often need to be revised into statements of achievement before they can be measured. This requires breaking standards down into component parts and linking them to specific actions.

Items were then tested with students in field studies in different states and among different demographic populations.

The prevalent standards, items, and student performance data were then scrutinized by a panel of three co-authors of the ISTE NETS-S 2007, and four technology instruction experts with classroom, district level, and state level experience. This panel, in conjunction with expert psychometricians, examined the data and made two determinations. First, they confirmed that 21<sup>st</sup> Century Skills Assessment does effectively measure student skills and knowledge in technology. Second, they determined where the bar for proficiency in technology literacy should lie for both the elementary and middle school national student populations. This determined the Proficiency Standard used in 21<sup>st</sup> century Skills Assessment. The panel then determined the cut-marks between below basic and basic, and between proficient and advanced.

**3. Was the purpose of the psychometric analysis to find distracter and poorly worded questions, and to determine the general difficulty of each question, along with how well it identified proficient students?**

**Answer:** The purpose of psychometric analysis is to ensure that an assessment's scores will be reliable and valid. Psychometric analysis is critical when test results will potentially be used for high stakes decisions. Some benefits are indeed to identify unusable questions, determine the general difficulty of each question, and determine how well each question identified proficient students, for example, the analysis shows how students that answered specific questions correctly or incorrectly did on the test as a whole. Item Response Theory was used for the analysis; you can find more information about IRT at:

[http://en.wikipedia.org/wiki/Item\\_response\\_theory](http://en.wikipedia.org/wiki/Item_response_theory) .

**4. Why do we produce a score between 100 and 500?**

**Answer:** The score begins at 100 to prevent confusion with percentage scores of 0 to 100%. It extends from 100 to 500 to prevent confusion resulting from trying to draw inaccurate relationships with other, unrelated assessments, and to provide 4 levels of proficiency.

**5. Why is the Proficiency Standard set at 300?**

The purpose of the scale score is to provide a readable, consistent method of comparing scores and showing proficiency that does not change between tests, even as the number of items required to be correct will change for each test based upon the psychometric analysis of the items.

300 was chosen to enhance readability of scores.

**6. Are the first right answers worth more points than the last right answers?**

**Answer:** Please note that the scaled scores indicate proficiency and are the **only** scores that are comparable to each other from one test to the next, not point values per item. 21<sup>st</sup> Century Skills Assessment uses a fixed cut mark at 300, or 200 points, to indicate proficiency. For each new version of the assessment (the pre and post test for an upcoming year) the combination of items is analyzed by psychometricians using Item Response Theory to determine the test characteristic curve which sets the new cutoff. The scaled passing score will always be 300; however, depending on the number of correct answers needed to obtain that score, the number of points each question is worth will vary from form to form, **as will the percentage of answers that need to be correct before a student is minimally proficient.**

This means that if the combination of items were shown to have a higher level of difficulty than before, less items will need to be correct to show. Or, if the analysis has determined that the items were less difficult than before, more items must be correct to achieve proficiency. Points per item are determined by psychometric analysis of each new test and usually vary between different cut-marks.

## 7. If each question is weighted equally, how much is each question worth?

**Answer:** Please note that the scaled scores indicate proficiency and are the **only** scores that are comparable to each other from one test to the next, not point values per item. While the number of points is calculated without numerical weighting, each new assessment has questions of varying difficulty. The combination of items for each new test is analyzed by psychometricians using Item Response Theory to determine the test characteristic curve which sets the new cutoff. The scaled passing score will always be 300; however, depending on the number of correct answers needed to obtain that score, the number of points each question is worth will vary from form to form, **as will the percentage of answers that need to be correct before a student is minimally proficient.**

This means that if the combination of items were shown to have a higher level of difficulty than before, less items will need to be correct to show proficiency. Or, if the analysis has determined that the items were less difficult than before, more items must be correct to achieve proficiency. Points per item are determined by psychometric analysis of each new test and usually vary between different cut-marks.