



ETS[®] Proficiency Profile

User's Guide

June 21, 2010

Copyright © 2010 by Educational Testing Service. All rights reserved.
ETS and the ETS logo are registered trademarks of Educational Testing Service.
No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage or retrieval system, without permission in writing from the copyright owners.

Table of Contents

ETS® Proficiency Profile	4
ETS Proficiency Profile Test Standard Form	5
Standard Form Structure	6
ETS Proficiency Profile Test Abbreviated Form.....	6
Abbreviated Form Structure.....	7
ETS Proficiency Profile Scaled Scores.....	7
Summary of Scaled Scores	8
Proficiency Classifications and Proficiency Level Statistics.....	9
Proficiency Levels.....	9
Skills Tested at Each Level	9
Proficiency Classifications on the Standard Form	11
Proficiency Classifications on the Abbreviated Form	12
Summary of Proficiency Classifications	13
Using the Standard Form versus the Abbreviated Form.....	14
Features of the Standard Form versus the Abbreviated Form.....	15
Flexible Administration	15
Paper-and-Pencil and Online Test Formats.....	15
Proctored versus Non-Proctored	16
How Colleges Use ETS Proficiency Profile	17
Growth Measurement.....	17
Trend Indicator	17
Comparisons with Other Institutions	17
Counseling Tool	17
Recruitment Aid	18
Comparative Data Guide.....	18
Guidelines for Test Use	19
Test Purpose	19
Selecting Students	19
Student Motivation.....	19
Limitations of Test Scores	19
Limitations of Comparative Data.....	20
Confidentiality of Score Data.....	20

continued

Student Sampling.....	20
Selecting a Simple Random Sample	21
Selecting a Stratified Random Sample.....	21
Selecting a Spaced Sample.....	22
Deciding How Many Students to Test.....	23
Reliability of Scaled Scores	23
Confidence Limits.....	25
What Are Confidence Limits	25
What Do Confidence Limits Indicate	26
What Is The “Total Population” and Why Does It Matter	26
How Strongly Do the Confidence Limits Depend on the Number of Students Tested	27
The ETS Proficiency Profile Essay.....	28

If you require additional assistance with the ETS Proficiency Profile or have questions that are not addressed in this *User’s Guide*, please call 1-800-745-0269, e-mail highered@ets.org, or visit the ETS Proficiency Profile section of the ETS website at www.ets.org/proficiencyprofile.

ETS® Proficiency Profile

The ETS® Proficiency Profile test was developed to assist in the assessment of the outcomes of general education programs in order to improve the quality of instruction and learning. It is a test of college-level skills in critical thinking, reading, writing and mathematics designed to measure the academic skills developed through general education courses, rather than the subject knowledge specifically taught in those courses. All of the subject knowledge required to answer each question is contained in the question itself or in the stimulus materials that accompany the question.

College-level reading questions measure students' ability to:

- interpret the meaning of key terms
- recognize the primary purpose of a passage
- recognize explicitly presented information
- make appropriate inferences
- recognize rhetorical devices

College-level writing questions measure students' ability to:

- recognize the most grammatically correct revision of a clause, sentence or group of sentences
- organize units of language for coherence and rhetorical effect
- recognize and reword figurative language
- organize elements of writing into larger units of meaning

Critical thinking questions measure students' ability to:

- distinguish between rhetoric and argumentation in a piece of nonfiction prose
- recognize assumptions
- recognize the best hypothesis to account for information presented
- infer and interpret a relationship between variables
- draw valid conclusions based on information presented

Mathematics questions measure students' ability to:

- recognize and interpret mathematical terms
- read and interpret tables and graphs
- evaluate formulas
- order and compare large and small numbers
- interpret ratios, proportions, and percentages
- read scientific measuring instruments
- recognize and use equivalent mathematical formulas or expressions

ETS Proficiency Profile is the successor to the Academic Profile, which was in use from 1990 until 2006. ETS Proficiency Profile is structurally identical to the version of the Academic Profile that was in use from 2001 to 2006, and ETS Proficiency Profile scores have been statistically equated for comparability to Academic Profile scores.

There are two versions of the ETS Proficiency Profile test:

- a two-hour **Standard form**, intended to provide information about individual students or groups of students
- a 40-minute **Abbreviated form**, intended to provide information about groups of 50 or more students

The ETS Proficiency Profile test yields two types of scores:

- scaled scores, which are numeric and norm-referenced
- proficiency classifications, which are categorical and criterion-referenced

ETS Proficiency Profile Test Standard Form

The Standard form of the ETS Proficiency Profile test is intended to provide information about individual students, as well as groups of students. It consists of 108 questions, divided into two sections of 54 questions each. The two sections may be administered either in a single two-hour testing session or in separate testing sessions of one hour each.

The Standard form includes:

- 27 questions testing critical thinking skills
- 27 questions testing reading skills
- 27 questions testing writing skills
- 27 questions testing mathematics skills

Most of the questions testing reading and critical thinking skills are in sets of two to four questions based on a common stimulus, such as a reading selection. Many of these sets include both reading and critical thinking questions. If the set includes only critical thinking questions, the stimulus may be something other than a reading passage, e.g., a picture or a graph.

The questions are sequenced in clusters:

- a group of reading and critical thinking questions
- a group of writing questions
- a group of mathematics questions
- a group of critical thinking questions, etc.

This sequence prevents all of the questions measuring a particular type of skill from appearing late in the test. Approximately half of the questions testing each type of skill appear in the first section of the test and half in the second section. The questions on the test are arranged in blocks of three to eight questions, with the questions in each block testing the same types of skill: reading and critical thinking skills, writing skills or math skills.

Most of the reading and critical thinking questions are based on information presented in a brief reading selection, picture, graph, etc. Each question measuring critical thinking or reading skills

is associated with a particular academic *context*: humanities, social sciences or natural sciences. The test includes nine critical thinking questions and nine reading questions in each of these three academic contexts. The table that follows illustrates this structure.

Standard Form Structure

	Critical Thinking	Reading	Writing	Mathematics
Humanities	9 questions	9 questions	27 questions	27 questions
Social Sciences	9 questions	9 questions		
Natural Sciences	9 questions	9 questions		

An institution administering the Standard form can obtain a Roster of Scaled Scores, showing the following scores for each individual student.

Total Score	Skills Subscores	Context-Based Subscores
	Critical Thinking	Humanities
	Reading	Social Sciences
	Writing	Mathematics
	Mathematics	

The institution can also obtain a Roster of Proficiency Categories, showing each student’s classifications as *proficient*, *marginal* or *not proficient* at each proficiency level in reading and critical thinking, writing and mathematics

The institution can also obtain a Summary of Scaled Scores and a Summary of Proficiency Classifications for any group of students tested, no matter how large or how small.

ETS Proficiency Profile Test Abbreviated Form

The Abbreviated form of the ETS Proficiency Profile test is not intended to provide information about individual students. It is intended to provide information about groups of at least 50 students. It can be administered in a single 40-minute testing session. The Abbreviated form is a partitioning of the Standard form into three smaller forms of 36 questions each, called “Abbreviated Form 1,” “Abbreviated Form 2” and “Abbreviated Form 3.” These three Abbreviated forms are packaged in alternating sequence, so that each of them is taken by one-third of the students. The 108 questions in the Standard form are assigned to the three Abbreviated forms to make each of them – insofar as possible – a miniature version of the Standard form.

Each of the three Abbreviated forms includes:

- nine questions measuring critical thinking skills
- nine questions measuring reading skills
- nine questions measuring writing skills
- nine questions measuring mathematics skills

On each Abbreviated form, the reading questions and the critical thinking questions represent all three of the academic contexts – humanities, social sciences and natural sciences – but the number of questions from each academic context in each subform can differ. The table that follows illustrates this structure.

Abbreviated Form Structure

	Critical Thinking	Reading	Writing	Mathematics
Humanities	3 questions	2, 3, or 4 questions	9 questions	9 questions
Social Sciences	3 questions	2, 3, or 4 questions		
Natural Sciences	3 questions	2, 3, or 4 questions		

The Abbreviated form provides too small a sample of each student’s performance to permit the reporting of individual scores (except for a total score based on all 27 questions). A student who takes the Abbreviated form is actually taking only one-third of the test, and the individual scores are not a reliable indication of the scores the student would have received on the full 108-question test. However, the larger the group of students, the more these inaccuracies will tend to cancel each other out.

An institution administering the Abbreviated form can obtain a Summary of Scaled Scores and a Summary of Proficiency Classifications for a group of at least 50 students. The institution can obtain a Roster of Scaled Scores, but the roster will contain only the students’ total scores; it will not include subscores. The institution cannot obtain a Roster of Proficiency Classifications for students who took the Abbreviated form.

ETS Proficiency Profile Scaled Scores

Eight scaled scores are reported for students taking the ETS Proficiency Profile test:

- a total score
- four skills subscores (critical thinking, reading, writing, mathematics)
- three context-based subscores (humanities, social sciences, natural sciences)

The total score is reported on a scale that extends from 400 to 500. The seven subscores are reported on a scale that extends from 100 to 130.

The scaled scores on each edition of the ETS Proficiency Profile test are computed in a way that adjusts for the difficulty of the questions to make them comparable to the corresponding scaled scores on other editions of ETS Proficiency Profile and Academic Profile. However, the different subscores are not necessarily comparable to each other. For example, a score of 125 in reading on any edition of the ETS Proficiency Profile test is comparable to a score of 125 in reading on any other edition of the ETS Proficiency Profile test or on Academic Profile. It is not comparable to a score of 125 in writing.

Each of the scaled scores is computed by a two-stage process:

1. The student’s “raw score” is computed by counting the number of questions the student answered correctly (there is no penalty for incorrect guessing).
2. A raw-to-scale conversion table converts the raw score to a scaled score.

There is a separate raw-to-scale conversion table for each of the eight scaled scores – the total score, the skill area subscores and the context-based subscores. The conversion tables for each edition of the ETS Proficiency Profile test were determined by a statistical process called score equating, to make them comparable to scaled scores on the other editions of the ETS Proficiency Profile test and on the Academic Profile. The equating process adjusts the raw-to-scale conversion to compensate for differences in the difficulty of the questions.

The scaled scores are norm-referenced, i.e., intended for comparing individual students with a group and for comparing a group of students with other groups of students. These comparisons can be entirely within the institution, or they can involve data from other institutions, available in the *Comparative Data Guide*.

Institutions using the Standard form receive a roster showing each individual student’s scaled scores – the total score, the skills subscores and the context-based subscores. All institutions using either the Standard form or the Abbreviated form receive a Summary of Scaled Scores report for each group – or “cohort” – of students tested. The same statistics are reported for the Standard form and for the Abbreviated form.

Summary of Scaled Scores

The Summary of Scaled Scores is a table of statistics describing the performance of a specified group of students taking the ETS Proficiency Profile test. If all students in the group took the Standard form, this report is available for a group of any size. If the group includes students who took the Abbreviated form, this report is available only for a group of at least 50 students. The reason for this restriction is that each student who takes the Abbreviated form actually takes only one-third of the full test. That student’s individual scores are not a reliable indication of the scores the student would have received on the full 108-question test. However, the larger the group of students, the more these inaccuracies will tend to cancel each other out.

The Summary of Scaled Scores includes the following statistics:

- The *mean score*. This statistic is simply the average score, computed by adding up the scores of all the students and dividing by the number of students.
- *95% confidence limits* for the mean score. These numbers define a probable range for the “true population mean,” i.e., the mean score that would result if all the students could somehow be tested with all possible forms of the ETS Proficiency Profile test.
- The *standard deviation*. This statistic describes the extent to which the students’ scores are spread widely throughout the score scale, rather than being bunched closely together.
- The *25th percentile* (also called the “first quartile”). This statistic is the score that separates the bottom 25 percent of the students from the rest of the group

- The *50th percentile* (also called the “median”). This statistic is the score that separates the lower half of the students from the upper half.
- The *75th percentile* (also called the “third quartile”). This statistic is the score that separates the top 25 percent of the students from the rest of the group.

These statistics are reported for the total score, the four skills subscores, and the three context-based subscores.

Proficiency Classifications and Proficiency Level Statistics

Proficiency Levels

The skills measured by the ETS Proficiency Profile test are grouped into three skill areas:

- reading and critical thinking
- writing
- mathematics

Within each of these three skill areas, the specific skills tested by the ETS Proficiency Profile test are classified into three *proficiency levels*, identified simply as **Level 1**, **Level 2** and **Level 3**. Each proficiency level is defined in terms of a set of specific competencies expected of students.

Skills Tested at Each Level

Reading and Critical Thinking

To be considered proficient at **Level 1**, a student should be able to:

- recognize factual material explicitly presented in a reading passage
- understand the meaning of particular words or phrases in the context of a reading passage

To be considered proficient at **Level 2**, a student should be able to:

- synthesize material from different sections of a passage
- recognize valid inferences derived from material in the passage
- identify accurate summaries of a passage or of significant sections of the passage
- understand and interpret figurative language
- discern the main idea, purpose, or focus of a passage or a significant portion of the passage

To be considered proficient at **Level 3**, a student should be able to:

- evaluate competing casual explanations
- evaluate hypotheses for consistency with known facts
- determine the relevance of information for evaluating an argument or conclusion
- determine whether an artistic interpretation is supported by evidence contained in a work
- recognize the salient features or themes in a work of art
- evaluate the appropriateness of procedures for investigating a question of causation
- evaluate data for consistency with known facts, hypotheses or methods.

Writing

To be considered proficient at **Level 1**, a student should be able to:

- recognize agreement among basic grammatical elements (e.g., nouns, verbs, pronouns and conjunctions)
- recognize appropriate transition words
- recognize incorrect word choice
- order sentences in a paragraph
- order elements in an outline

To be considered proficient at **Level 2**, a student should be able to:

- incorporate new material into a passage
- recognize agreement among basic grammatical elements (e.g., nouns, verbs, pronouns and conjunctions) when these elements are complicated by intervening words or phrases
- combines simple clauses into single, more complex combinations
- recast existing sentences into new syntactic combinations.

To be considered proficient at **Level 3**, a student should be able to:

- discriminate between appropriate and inappropriate use of parallelism
- discriminate between appropriate and inappropriate use of idiomatic language
- recognize redundancy
- discriminate between correct and incorrect constructions
- recognize the most effective revision of a sentence

Mathematics

To be considered proficient at **Level 1**, a student should be able to:

- solve word problems that would most likely be solved by arithmetic and do not involve conversion of units or proportionality (These problems can be multi-step if the steps are repeated rather than embedded.)
- solve problems involving the informal properties of numbers and operations, often involving the Number Line, including positive and negative numbers, whole numbers and fractions (including conversions of common fractions to percent, such as converting $\frac{1}{4}$ to 25%)
- solve problems requiring a general understanding of square roots and the squares of numbers
- solve a simple equation or substitute numbers into an algebraic expression
- find information from a graph (This task may involve finding a specified piece of information in a graph that also contains other information.)

To be considered proficient at **Level 2**, a student should be able to:

- solve arithmetic problems with some complications, such as complex wording, maximizing or minimizing and embedded ratios (These problems include algebra problems that can be solved by arithmetic [the answer choices are numeric].)
- simplify algebraic expressions, perform basic translations and draw conclusions from algebraic equations and inequalities (These tasks are more complicated than solving a simple equation, though they may be approached arithmetically by substituting numbers.)

- interpret a trend represented in a graph, or choose a graph that reflects a trend
- Solve problems involving sets (The problems would have numeric answer choices.)

To be considered proficient at **Level 3**, a student should be able to:

- solve word problems that would be unlikely to be solved by arithmetic; the answer choices are either algebraic expressions or are numbers that do not lend themselves to back-solving
- solve problems involving difficult arithmetic concepts such as exponents and roots other than squares and square roots and percent of increase or decrease
- generalize about numbers, e.g., identify the values of (x) for which an expression increases as (x) increases
- solve problems requiring an understanding of the properties of integers, rational numbers, etc.
- interpret a graph in which the trends are to be expressed algebraically or in which one of the following is involved: exponents and roots other than squares and square roots, percent of increase or decrease
- solve problems requiring insight or logical reasoning.

Proficiency Classifications on the Standard Form

The skills measured by the ETS Proficiency Profile test are grouped into three skill areas:

1. reading and critical thinking
2. writing
3. mathematics

Within each of these three skill areas, the specific skills tested are grouped into three proficiency levels:

- Level 1
- Level 2
- Level 3

The specific competencies have been grouped into proficiency levels in such a way that most students who are proficient at the higher-level skills will also be proficient at the lower-level skills in the same skill area. For example, most of the students who are proficient at the Level 3 writing skills will also be proficient at the Level 2 writing skills. Each student taking the Standard form of the ETS Proficiency Profile test is classified as *proficient*, *marginal* or *not proficient* at each proficiency level of each skill area. A student classified as marginal is one whose test results do not provide clear enough evidence to classify the student as proficient or not proficient.

Each question on the ETS Proficiency Profile test is associated with a particular proficiency level in a particular skill area (e.g., Level 2 in writing). In writing and in mathematics, each proficiency classification reported for a student depends on the student's responses to the questions from all three proficiency levels. However, the questions are weighted unequally, with the questions at the relevant proficiency level taking the greatest weight. For example, in classifying a student as proficient, marginal or not proficient at Level 2 in mathematics, all 27

mathematics questions enter into the formula, but the Level 2 mathematics questions are given more weight than the Level 1 questions or the Level 3 questions.

The procedure is slightly different for reading and critical thinking. Proficiency Levels 1 and 2 consist of reading skills, but Level 3 consists of critical thinking skills, which go beyond reading skills. The proficiency classifications for Level 1 and Level 2 reading are based on the 27 reading questions; the 27 critical thinking questions are not included. Again, the questions are weighted unequally, with the questions at the relevant proficiency level taking the greater weight. The proficiency classification for Level 3 critical thinking is based only on the 27 critical thinking questions.

The cut points for the proficiency classifications on the ETS Proficiency Profile test were determined by a statistical procedure designed to make them comparable – as nearly as possible – to those on the Academic Profile, adjusting for any differences in the difficulty of the questions. Because items are weighted differently for each skill area for each test form (there are several forms for paper and computer administrations), there is no simple way to translate a specific number right (NR) score to a proficiency level.

All institutions receive a “Summary of Proficiency Classifications” report for each group of students – or “cohort” – tested. The report shows the percentages of the students classified as proficient, marginal and not proficient, at each proficiency level in each skill area.

Institutions using the Standard form of the ETS Proficiency Profile test also receive a roster showing these proficiency classifications for each individual student (P = proficient; M = marginal; N = not proficient).

Proficiency Classifications on the Abbreviated Form

A student taking the Abbreviated form of the ETS Proficiency Profile test answers only three questions at each proficiency level in writing and in mathematics, and only four or five at each level in reading. Therefore, the Abbreviated form does not provide sufficient information for proficiency classifications of individual students. However, over a large enough group of students, the inaccuracies tend to average out, making it possible to estimate the percentage of the group who would be classified as proficient, marginal and not proficient, if the students had taken the Standard form.

The percentages of students at each proficiency level in reading, writing and mathematics are estimated by computing a proficiency classification for each student. Each classification makes use of the student’s responses to all the questions in the skill area but weights the questions unequally. The individual classifications are not reported but are used to compute an estimated percentage for the group. The cut points for the classifications were chosen by a statistical procedure designed to make the group percentages on the Abbreviated form comparable – as nearly as possible – to those on the Standard form.

All institutions receive a “Summary of Proficiency Classifications” report for each group of students whose answer sheets are submitted as a separate batch. The report shows the percentages of the students classified as proficient, marginal and not proficient, at each proficiency level in each skill area.

Summary of Proficiency Classifications

The skills measured by the ETS Proficiency Profile test are grouped into three skill areas:

1. reading and critical thinking
2. writing
3. mathematics

Within each of these three skill areas, the specific skills tested are grouped into three proficiency levels:

- Level 1
- Level 2
- Level 3

The specific competencies have been grouped into proficiency levels in such a way that most students who are proficient at the higher-level skills will also be proficient at the lower-level skills in the same skill area. For example, most of the students who are proficient at the Level 3 writing skills will also be proficient at the Level 2 writing skills. At each proficiency level in each skill area, a student can be classified as proficient, marginal, or not proficient.

The Summary of Proficiency Classifications describes the performance of a specified group of students taking the ETS Proficiency Profile test. It includes a table and a graph showing the percentages of the group classified as proficient, marginal, and not proficient, at each of the three proficiency levels in each of the three skill areas. If all the students in the group took the Standard form of the ETS Proficiency Profile test, these percentages refer to the classifications reported for the individual students. If some or all of the students took the Abbreviated form, the percentages must be interpreted as estimates for the group.

If all of the students in the group took the Standard form, this report is available for a group of any size. If the group includes students who took the Abbreviated form, this report is available only for a group of at least 50 students. The reason for this restriction is that each student who takes the Abbreviated form actually takes only one-third of the test. That student’s individual proficiency classifications are not a reliable indication of the classifications the student would have received on the full 108-question test. However, the larger the group of students, the more these inaccuracies will tend to cancel each other out.

Using the Standard Form versus the Abbreviated Form

Some institutions may choose to use the ETS Proficiency Profile test to assess the skills of individual students. Others may use it only to characterize the skills of groups of students, for example, an incoming freshman class or a graduating senior class. In selecting a test to assess general education outcomes, an institution should begin by considering its purpose in wanting to test. How will the test results be used? Is it important to assess the skills of each individual student, or is it sufficient to assess a class of students as a group?

The Abbreviated form requires only 40 minutes of testing time, but it provides only group information – a set of statistical reports for each group of students – or “cohort” – tested plus some additional information on subgroups of the students determined from the demographic data. The Abbreviated form is constructed by dividing the Standard form into three parts, and packaging them in alternating sequence (A, B, C, A, B, C, A, B, etc.) so that each part is taken by one-third of the students. The alternating sequence makes it likely that the groups taking the three parts will be similar, particularly if the number of students is fairly large. This sampling technique (sometimes called “matrix sampling”) makes it possible to obtain reliable information about the group even when no individual student answers enough questions to provide reliable individual subscores.

The Standard form requires two hours of testing time, but it provides scores and proficiency classifications for individual students, in addition to the group information provided by the Abbreviated form. The individual information can be used in advising students and in making decisions about individual students.

Data from both the Standard and Abbreviated forms can also be used to compute statistics that are not routinely provided, for the full group or for subgroups of the students tested. With the eight scaled scores and the nine proficiency classifications for each student tested, it is possible to aggregate and disaggregate the data in ways that expand the institution’s information base. Data for a particular cohort can be downloaded in (Microsoft® Excel® format) for additional analysis. *However, because subscores and proficiency classifications on the Abbreviated form should not be considered adequately reliable at the individual student level, the data downloaded from the Abbreviated form administration will not contain student identifiers such as names or student ID numbers in order to prevent these scores from being used to make high-stakes decisions about individual students.* So, while aggregation and disaggregation of data is possible with both the Standard and Abbreviated forms, without student identifiers it is not possible to compare data from the Abbreviated form administrations to other academic indicators for individual students such as class grades, GPAs or other test scores.

An institution’s decision to use the Standard form versus the Abbreviated form of the ETS Proficiency Profile test will depend mainly on the institution’s purpose in testing. An institution must decide whether to give up the individual scores and proficiency classifications provided by the Standard form in exchange for the reduction in testing time offered by the Abbreviated form. As the needs and priorities of a particular institution evolve, the institution can consider switching from one form to another based on the different benefits these different forms offer. Because the Abbreviated form is derived from the Standard form, the Abbreviated form is also

statistically equated to the Standard form – making the scores on each form fully comparable to scores on the other form. The two forms can be used interchangeably and scores from each can be compared with full confidence that they mean the same thing and can be interpreted the same way. This also implies that aggregation of data from both the Standard and Abbreviated forms is possible. *However, because subscores and proficiency classifications are not considered adequately reliable at the student level, reports derived from a combination of both Standard form and Abbreviated form test takers only include summary data.*

Features of the Standard Form versus the Abbreviated Form

Features	Standard Form	Abbreviated Form
Test administration	At convenience of institution	At convenience of institution
Testing time	2 hours (can be administered in two separate sessions)	40 minutes
Number of questions	108	36, in each of three spiraled forms
Norm-referenced scores (scaled scores)	Total score and seven subscores reported for each student	Total score reported for each student
	Group statistics reported for total score and seven subscores	Group statistics reported for total score and seven subscores
Criterion-referenced proficiency information	Proficiency classifications reported for individual students Group percentage proficient, marginal, and not proficient reported for each proficiency level in each skill area	No information reported for individual students (<i>except in data download without student identifiers</i>) Group percentage proficient, marginal, and not proficient reported for each proficiency level in each skill area
Demographic data	Group percentages reported	Group percentages reported
Demographic subgroup statistics	Statistics reported	Statistics reported
Locally written questions	Up to 50 multiple-choice questions	Up to 50 multiple-choice questions

Flexible Administration

Paper-and-Pencil and Online Test Formats

The ETS Proficiency Profile test is available in both the traditional paper-and-pencil format and in an Internet-based online format. The paper-and-pencil format is intended to be administered in a proctored environment with strict adherence to the administration procedures detailed in the

ETS Proficiency Profile *Supervisor's Manual*. The online version is delivered using a secure Web browser. While taking the online version of the test, the secure browser prevents students from using their PCs for other purposes which might constitute a testing impropriety. This and other security measures make the online version of ETS Proficiency Profile test suitable for both proctored and non-proctored administrations. Procedures for administering the online test are detailed in the ETS Proficiency Profile *Institution Administrator Manual* and ETS Proficiency Profile *Proctor Administrator Manual*. The paper-and-pencil and online versions of the ETS Proficiency Profile test are statistically equated to each other – making the scores on each version fully comparable to scores on the other version. Depending on an institution's needs and technological infrastructure, the paper-and-pencil and online versions can be used interchangeably or in combination with each other. Scores from each version can be compared with full confidence that they mean the same thing and can be interpreted the same way. Aggregation of data from both the paper-and-pencil and online versions is possible.

Proctored versus Non-proctored

There are also options for non-proctored versions of the ETS Proficiency Profile test. The non-proctored versions allow institutions to reach their distance learning students and help facilitate sampling of off-campus students. The non-proctored versions of the ETS Proficiency Profile test are designed to be administered in an unproctored environment at the student's convenience. The institution should employ any reasonable policies and measures that it would ordinarily apply to its other educational activities – such as online coursework and evaluations, for example – to help ensure test security and student integrity. To this end, institutions should also consider the duration of their testing window and discourage the discussion of test questions among students to minimize exposure of the test questions.

Like the proctored versions of the ETS Proficiency Profile tests, the non-proctored versions are also available in Standard and Abbreviated forms. However, to maintain test security and integrity, the unproctored versions are comprised of multiple forms, each containing different sets of test questions. These multiple forms are all statistically equated to each other so that scores from any form of the test are 100% comparable to scores from other forms. Test security and integrity are ensured by randomly administering these multiple test forms. While random matrix sampling helps ensure test security, the ETS Proficiency Profile test also employs technological advances to aid in the prevention of testing improprieties. The unproctored ETS Proficiency Profile test is administered in a “secure browser” in which students cannot perform other functions while taking the ETS Proficiency Profile test. All keyboard functions that are not essential to the administration of the ETS Proficiency Profile test are disabled, and any attempt to access these keys or functions will result in the student being automatically ejected from the testing session. However, ETS does not expect that students would be motivated to attempt a testing impropriety on what is typically considered a “low-stakes exam,” given that the test rarely counts for or against individual students in any meaningful way.

How Colleges Use ETS Proficiency Profile

Growth Measurement

The ETS Proficiency Profile test can be used to assess students' growth in reading, writing, critical thinking and mathematics skills by testing the same students at different times in their educational careers, e.g., as incoming freshmen and as rising juniors or as graduating seniors. However, a group of incoming freshmen will include many students who will not be available to be tested as rising juniors or as graduating seniors. Therefore, this use of the ETS Proficiency Profile test requires individual scores. The Abbreviated form provides individual total scores, but no individual subscores or proficiency classifications. To measure growth in the specific types of skills reflected in the subscores and proficiency classifications, the institution must use the Standard form. The scaled scores are generally better than the proficiency classifications for this purpose, because some students (particularly those who are quite weak initially) may make gains that are not large enough to change their proficiency classifications. These gains will be reflected in their scaled scores.

Trend Indicator

Some colleges use the ETS Proficiency Profile test to look for year-to-year changes in the skill levels of their incoming freshmen, their rising juniors, or their graduating seniors. Both the scaled scores and the proficiency classifications can be used for this purpose, although the scaled scores may reflect differences that are not apparent in the proficiency classifications. In these year-to-year comparisons, the groups being compared consist of different students; there is no need to match students' scores from one testing with their scores from another testing. Either the Standard form or the Abbreviated form (if the groups are large enough) will provide the necessary data.

Comparisons with Other Institutions

Some colleges use the ETS Proficiency Profile test to determine how their students' skills compare with the skills of students at similar institutions. It is important that these comparisons involve students at the same point in their educational careers (entering freshmen, etc.). The information about the performance of students at other institutions can come from either of two sources: the *Comparative Data*. The norm-referenced scaled scores are particularly useful for this purpose, although the criterion-referenced proficiency classifications can also be used. Either the Standard form or the Abbreviated form (if the groups are large enough) will provide the necessary data.

Counseling Tool

Some colleges test their students to identify those whose reading, writing or mathematics skills need improvement, so that they can advise these students to take courses specifically aimed at improving those skills. Some colleges test their students twice, as entering freshmen and as rising juniors, to identify the skills that the students need to improve as freshmen and the skills that they still need to improve as rising juniors. The proficiency classifications are likely to be especially useful for this purpose, because each proficiency level is associated with a particular set of specific skills. This use of the ETS Proficiency Profile test requires the Standard form.

Recruitment Aid

Some colleges administer the Standard form of the ETS Proficiency Profile test to the same students as incoming freshmen and as rising juniors or as graduating seniors. They use this information to identify those students whose skills improved substantially. They then look for information in the students' educational background – information available at the time of admission – that distinguishes these students from those whose skills did not improve. This information can help the college focus its recruitment efforts on students who are likely to benefit from the college's instructional program. Both the scaled scores and the proficiency classifications are useful for this purpose. This use of the ETS Proficiency Profile test requires the Standard form.

Comparative Data Guide

The *Comparative Data Guide*, available in the ETS Proficiency Profile section of the ETS website at www.ets.org, is a collection of statistical tables describing the performance of various groups of students taking the Academic Profile from 2001 to 2006. The scores and proficiency classifications on the ETS Proficiency Profile test have been statistically equated for to make them comparable to the corresponding scores and classifications on the Academic Profile. The *Comparative Data Guide* contains statistics computed for groups of students classified by the point in their academic career at which they were tested and by the Carnegie classification of their colleges. The guide includes statistics for the following classes of students:

- Entering Freshmen (no credit hours completed)
- Freshmen (fewer than 30 semester hours or fewer than 45 quarter hours completed)
- Sophomores (30 to 60 semester hours or 45 to 90 quarter hours completed)
- Juniors (61 to 90 semester hours or 91 to 145 quarter hours completed)
- Seniors (more than 90 semester hours or more than 145 quarter hours completed)

All students tested, grouped by Carnegie classification:

- Doctoral/Research Universities I and II
- Master's (Comprehensive) Colleges and Universities I and II
- Baccalaureate (Liberal Arts) Colleges I and II
- Associate's Colleges
- Specialized Institutions
- All institutions

For each of these groups of students (e.g., entering freshmen at Doctoral/Research Universities), the *Guide* provides the following types of information:

- The number of students tested at each institution
- The distribution of the institutions' mean scaled scores (total scores and subscores) and the mean and standard deviation of this distribution
- The distribution of individual students' scaled scores – total scores and subscores – the percent scoring below each scaled score, and the mean and standard deviation of this distribution.
- The percentage of the students classified as proficient, marginal and not proficient at each proficiency level in reading and critical thinking, in writing, and in mathematics
- Demographic statistics describing the group of students

Guidelines for Test Use

Institutions planning to use the ETS Proficiency Profile test should be aware of its possibilities **and** its limitations. The following guidelines are provided to assist institutions in using the ETS Proficiency Profile test appropriately.

Test Purpose

Review examples of the test items and the list of skills it measures to verify that the skills that the test measures are those that the institution seeks to measure. Examine samples of the score reports and statistical reports, to verify that they include the information that the institution needs. If the institution needs information about individual students, make sure to use the Standard form, not the Abbreviated form.

Selecting Students

If the purpose of testing is to make inferences about the performance of groups of students, it is important to test an adequate number of students **from each of those groups**, selected in such a way that **the students tested from each group are representative of the group as a whole**. The best way to accomplish this is to test all of the students. If the institution tests only a sample of the students (see “Sampling of Students”), it is important that the sample include an adequate number of students from each group about which the institution wants information, selected in a way that will permit the results to be generalized to the group as a whole. It is particularly important not to limit the testing to students who volunteer to be tested, unless the institution wants information that applies only to those students.

Student Motivation

Student motivation in assessment testing is a serious concern. If the students are not motivated to do well on the test, their test scores will not reflect their actual skill levels.

Limitations of Test Scores

A test contains only a sample of the tasks that students are expected to be able to do. On another sample of tasks designed to measure the same skills, the students might perform somewhat differently. Information provided on the score reports (see “Confidence Limits”) and elsewhere in this *User’s Guide* (see “Reliability of Scaled Scores”) enables the user of the scores to determine how much the scores could be expected to differ if a different set of tasks were used.

The reliability of the individual scores of students taking the Standard form – particularly, the skill area scores and the proficiency classifications – should be adequate for counseling purposes and for the identification of students with problems in particular skill areas. **These scores are not reliable enough to use as the basis for high-stakes decisions about individual students.**

The ETS Proficiency Profile test measures a specific collection of skills. It does not and cannot measure all the educational outcomes of interest to institutions of higher learning. When the ETS Proficiency Profile test is used to evaluate an institution or any of its programs, it should be used in conjunction with other information. It should never be used as the sole means for evaluating the effectiveness of an institution or the educational progress of the students.

Limitations of Comparative Data

The data in the *Comparative Data Guide* are drawn entirely from institutions that use the ETS Proficiency Profile test. Within any category of institutions, those that use the ETS Proficiency Profile test are not likely to be representative of all institutions in that category. In addition, the numbers of students tested and the sampling procedures vary from one institution to another, and it is impossible to verify that the students tested at each institution are representative of all the institution's students at the relevant class level (freshman, sophomore, etc.).

Confidentiality of Score Data

ETS treats all score data for individuals and for institutions as confidential. Individual data are released only to the institution of the students tested. Identifiable institutional data are released only to the institution providing the data, unless the institution gives written permission to release the information to others. Institutions using the ETS Proficiency Profile test should adopt a similar policy for the data from their individual students.

Student Sampling

If an institution's purpose in using the ETS Proficiency Profile test is only to assess the skills of a group, not of individual students, and if that group includes too many students to test them all, it may be adequate to test only a sample of the students – but **only** if the sample is selected in such a way that the students taking the test **will not differ systematically from those who are not tested**.

Statisticians involved in sampling use the term “population” to refer to the group of people that they want their statistics to describe. If you are testing with the ETS Proficiency Profile test, the population is the group of students you want the ETS Proficiency Profile statistics to describe – for example, an entire entering freshman class.

Selecting and testing a sample of students that will not differ systematically from the full population is often difficult. Students who volunteer to take a test – and who then actually show up and take the test – almost always differ systematically from students who do not. Students enrolled in particular courses tend to differ systematically from students not enrolled in those courses. Testing the students whose schedules make the testing convenient is likely to give you a sample that differs systematically from the full population. **The safest way to avoid these systematic differences is to test all of the students.** However, it is often possible to select a sample of students that will adequately represent the full population.

Before you decide to test only a sample of the students, you should ask yourself several questions. If you do not test all the students, can you be sure that the students you select will actually show up and take the test? Are you interested only in describing the skills of the whole population, or do you also want statistics for subgroups of the population (for example, resident students and commuting students, or students in different degree programs)? If you are interested in subgroups, will your sample include enough students from each of those subgroups? The smaller the sample of students from a subgroup, the less likely that the statistics will generalize

to the entire subgroup. Also, if you use the Abbreviated form, you will not be able to obtain a Summary of Scaled Scores or a Summary of Proficiency Classifications for a group of fewer than 50 students.

If you decide to test only a sample of the students, it is important to select them by a sampling procedure that makes it unlikely that the sample will be very different from the full population. Here are some sampling procedures you can use to select a sample that will be representative of the full population. However, it is not enough simply to select the students and schedule them for testing. If you want your sample of students to be representative of the population, you must make sure that all or nearly all of the students you select actually take the test. The greater the proportion of the students in your sample who do not take the test, the more your sample is likely to differ from the population.

Selecting a Simple Random Sample

A simple random sample of students can differ from the population, but if the sample is large, it is likely to resemble the population fairly closely. To select a simple random sample, first assign a number to each student in the population. You can assign the numbers in any convenient way. Next, decide how many students you want to schedule for testing. This number is your sample size. Then use your computer's random number generator, or a table of random numbers, to select students by number, one at a time, until you have selected as many students as you plan to schedule for testing.

Selecting a Stratified Random Sample

To make your sample resemble the population more closely, you can divide the full population into groups that are likely to perform differently on the test. You can then select a simple random sample separately from each of those groups. Each of those groups is called a “stratum” (plural: “strata”) and the process is called “stratification.”

To select a stratified random sample, first look at the information that you have about each student in the population, and choose one or more characteristics likely to be related to students' performance on the test. (To keep the procedure reasonably simple, keep the number of separate characteristics small.) For example, suppose you anticipate a systematic difference in test performance between day students and evening students. In that case, the proportion of day students to evening students should be the same in your sample as it is in the population. You might also want to make sure that students in certain academic fields (e.g., liberal arts, business, health sciences) are represented in the same proportion in your sample as in the population.

Next, divide the population into “strata” representing the possible combinations of the student characteristics you decided to use. Continuing with the previous example, you would have a stratum for day students in liberal arts, another for evening students in liberal arts, another for day students in business, etc. Make sure that each student in the population is included in **one and only one** stratum. If some of the strata have very few students, you can combine them. For example, you might create a single stratum for evening students in all academic areas except business and health science.

Next, decide how many students you want to schedule for testing. This number is your total sample size. Divide your total sample size by the total number of students in the population, to determine what percentage of the population you will schedule for testing. Then select that percentage of the students in each stratum, by taking a simple random sample from each stratum. In the example above, if you intended to test 40 percent of the students in the population, you would select 40 percent of the students in each stratum: 40 percent of the day students in liberal arts, 40 percent of the evening students in liberal arts, 40 percent of the day students in business, etc.

Selecting a Spaced Sample

Another way to make your sample resemble the population is to use “spaced sampling.” A spaced sample consists of every fifth student, or every eighth student, etc., on a sorted list of all the students. The resulting sample will closely represent the population **with respect to those characteristics used to sort the list**.

To use this technique, first make a list of all the students in the population. Then sort the list according to the characteristics on which it is most important that the sample of students tested represent the full population. These characteristics might include day vs. evening, academic area, demographic characteristics such as gender and ethnicity, and possibly even academic performance information such as grade-point average (GPA). If you use GPA, it may be wise to sort on GPA **within** each academic area, to allow for possible differences in grading standards. After you have sorted the list, number the students according to their positions on the sorted list, from 1 to the population size.

Next, decide how many students you want to schedule for testing. This number is your sample size. Divide the population size by the sample size. The resulting number is your “sampling interval.” For example, if there are 1,200 students in the population and you intend to test 200, your sampling interval is 6; you will select every sixth student on the list.

Next, choose **at random** (using your computer’s random number generator or a random number table) a number from 1 to the sampling interval. In the above example, you would choose at random a number from 1 to 6. This number is your “random start” and it identifies the first person from the list who you will select into the sample. After that, select students at regular intervals equal to the sampling interval. Continuing with the example, suppose your random start is 4 and the sampling interval is 6. The first student selected into the sample would be Student #4 on the sorted list. Since $4 + 6 = 10$, the next student selected would be Student #10. Since $10 + 6 = 16$, the next student selected would be Student #16. And so on, until you have reached the end of the list.

One limitation of spaced sampling is that you cannot easily select a sample of any specified size. For example, if there were 1,100 students in the population, you could not easily select a sample of 200 students. A sampling interval of 5 would give you a sample of 220 students; a sampling interval of 6 would give you a sample of 183 students. If you wanted exactly 200 students in your sample, you could first select a spaced sample of 183 students by using a sampling interval of 6. You could then select another 17 of the 917 students not yet selected, by taking a second spaced sample, using a sampling interval of 54. (This second sample would require another random start, a randomly generated number from 1 to 54.)

Deciding How Many Students to Test

In deciding how many students to test, it is important to remember that the smaller the sample of students tested, the more the statistics computed from the sample are likely to differ substantially from what they would be if all the students in the population were tested. Also, the smaller the sample, the more important it is to test each individual student in the sample. One student missing from a sample of 100 students will affect the results much more than one student missing from a sample of 500.

The following table shows the sample sizes necessary for two different levels of accuracy in using the percent proficient in a sample of students taking the Abbreviated form, as an estimate of the percent proficient in the full population. The table answers the question, “If exactly 50 percent of the students in the population are proficient, how many students would the sample have to include, for a probability of .95 that the sample percentage will be within ten percentage points of that number (40% to 60%)? Within five percentage points (45% to 55%)?”¹

**Required Sample Size (Abbreviated Form), if 50%
of the Students in the Population are Proficient**

Population Size	For a .95 Probability that the Percent Proficient in the Sample Will Be	
	40% to 60%	45% to 55%
1000	94	342
500	91	307
300	87	271
200	84	*
100	74	*

*This level of accuracy is not possible with the Abbreviated form of the ETS Proficiency Profile test.

Reliability of Scaled Scores

In educational testing, reliability means consistency. A student’s score on a test is a very specific piece of information. It describes the student’s performance on a particular set of test questions. But the questions on the test the student took are not the only questions that could have been used to measure those skills. If different questions had been used to measure those skills, how different would the student’s score have been? It is not possible to answer this question for an individual student, unless the student actually took two different editions of the test containing

¹ These calculations assume simple random sampling. The analysis was based on the proficiency classifications for Reading, Level 1, in a group of 705 students at 6 colleges that participated in the ETS Proficiency Profile pilot study. The overall percent proficient in this combined group was 50%; the percentages at the six individual colleges were 43%, 47%, 50%, 52%, 55%, and 58%.

different questions – and even then, the answer would apply only to those two editions. But it is possible to estimate statistics that indicate, for a large group of students, how closely their scores on two different editions of the test would agree.

The statistics most commonly used to describe the reliability of test scores are the *reliability coefficient* and the *standard error of measurement*.² The reliability coefficient is an estimate of the correlation between students' scores under two different conditions of testing – most often, two different editions of the test. The reliability coefficient describes, on a scale of .00 to 1.00, the extent to which the students who scored high under one condition (e.g., the set of questions they actually took) would also score high under the other condition (e.g., another set of questions measuring the same skills).³ The reliability coefficient can vary substantially from one group of students to another. It tends to be larger for groups of students whose scores differ greatly from each other and smaller for groups of students whose scores are similar.

The standard error of measurement (SEM) is expressed in the same units as the test scores (e.g., scaled-score points). The SEM indicates how much students' scores tend to differ, on the average, from the scores they would get by taking many different editions of the test and averaging the results.⁴ The more reliable the students' scores are, the closer their scores are to the scores they would get under this hypothetical averaging process, and the smaller the SEM will be. In a large group of students, about two-thirds of the students will have scores within one SEM of the scores they would get under this averaging process. Unlike the reliability coefficient, the SEM tends to be similar for different groups of students taking the same test.

The following table shows the reliability coefficient and the SEM of the scores of the 2,618 students participating in the ETS Proficiency Profile pilot study conducted in 2005.⁵

² “Error of measurement” is the term that people in the testing profession use to refer to all the things that can make test scores less than perfectly reliable. It does not imply that someone has made a mistake in administering or scoring the test. In the language of psychometrics, the “error of measurement” in a student’s test score is the difference between that score and the score the student would have earned if the test scores were perfectly reliable. The standard error of measurement of the scores of a group of students is the standard deviation of the distribution of the errors of measurement in their scores.

³ For example, suppose the test scores of a very large group of students have a reliability coefficient of .90. If the students with scores one standard deviation above the group mean were to take another edition of the test, their average score on that edition would be 0.9 standard deviations above the group mean.

⁴ Psychometricians refer to this hypothetical average score as the student’s “true score.”

⁵ Each student took only one of the four different editions of ETS Proficiency Profile that were included in the pilot study. The reliability statistics were averaged over these four editions.

Reliability Statistics for Individually Reported Scaled Scores

	Number of Questions	Reliability Coefficient	Score Scale	Standard Error of Measurement
Standard Form				
Critical Thinking	27	0.78	100 to 130	3.00
Reading	27	0.80	100 to 130	3.28
Writing	27	0.81	100 to 130	2.34
Mathematics	27	0.84	100 to 130	2.46
Humanities	18	0.68	100 to 130	3.73
Social Sciences	18	0.73	100 to 130	3.37
Natural Sciences	18	0.75	100 to 130	3.11
Total Score	108	0.91	400 to 500	6.04
Abbreviated Form				
Total Score	36	0.77	400 to 500	9.64

Technical note: The reliability statistics for the writing and mathematics subscores were estimated by the commonly used “alpha” reliability formula, which is based on the statistical relationships among all the individual questions. The reliability statistics for the scores that include reading or critical thinking questions were estimated with the “Kristof” three-part procedure. Most of the reading and critical thinking questions in the ETS Proficiency Profile test are contained in “item sets” consisting of two or more questions based on a common reading paragraph or other stimulus material. The “alpha” reliability formula, applied in the usual way (i.e., treating each individual question as a separate observation), will tend to overestimate the reliability of scores on a test in which many of the questions are grouped into “item sets.” The Kristof procedure was applied by dividing the Standard form into three closely similar parts, keeping the item sets intact, i.e., the three abbreviated forms.

Confidence Limits

What Are Confidence Limits?

A student’s scaled score on the ETS Proficiency Profile test – either the total score or one of the subscores – is computed from the questions administered to that student. But the questions on a particular form of the ETS Proficiency Profile test – even those on the Standard form – are only a small sample of the thousands of possible questions that could have been used to measure those same skills. If the questions had been different, that student’s scaled score might well have been different. The same is true of the scores of the other students in the group. If the questions had been different, some students might have had higher scores; some might have had lower scores. These differences in individual students’ scores would tend to average out – but not completely. They would cause a difference in the group mean score.

In some cases, the students taking the test are the only students the test user is interested in generalizing to. But in other cases, they are not. The students taking the test may have been sampled from some larger group of students. Even if they were not, the test user may be interested in generalizing beyond the specific students tested, to draw conclusions about a very

large group of students like the students actually tested. The group of students to which the test user would like to generalize is called the *population*.

The mean score that would be most useful, if it could somehow be determined, would be the mean score that would result if every student in the population could be tested with all of the thousands of possible questions that could have been used to test the skills that the ETS Proficiency Profile test measures. This hypothetical mean score is called the “true population mean.” It is not possible to compute the true population mean, but it is possible to determine limits for it, in such a way that the test user can be highly confident that the true population mean lies between these limits. These limits are called *confidence limits*.

What Do the Confidence Limits Indicate?

The confidence limits reported for each ETS Proficiency Profile scaled score mean are “95% confidence limits” for the true population mean. That is, they are computed by a procedure that has a 95 percent probability of producing limits that will surround the true population mean. If the procedure were repeated with many different sets of data, 95 percent of the resulting pairs of limits would surround the true population mean. In that sense, the score user can be “95% confident” that the true population mean lies between the confidence limits. The closer together the confidence limits are, the more precisely the test user can locate the true population mean. The mean score of the students actually tested is always between the confidence limits, and if the confidence limits are close together, the test user can be confident that the true population mean is close to the mean score of the students actually tested.

What Is The “Total Population” and Why Does It Matter?

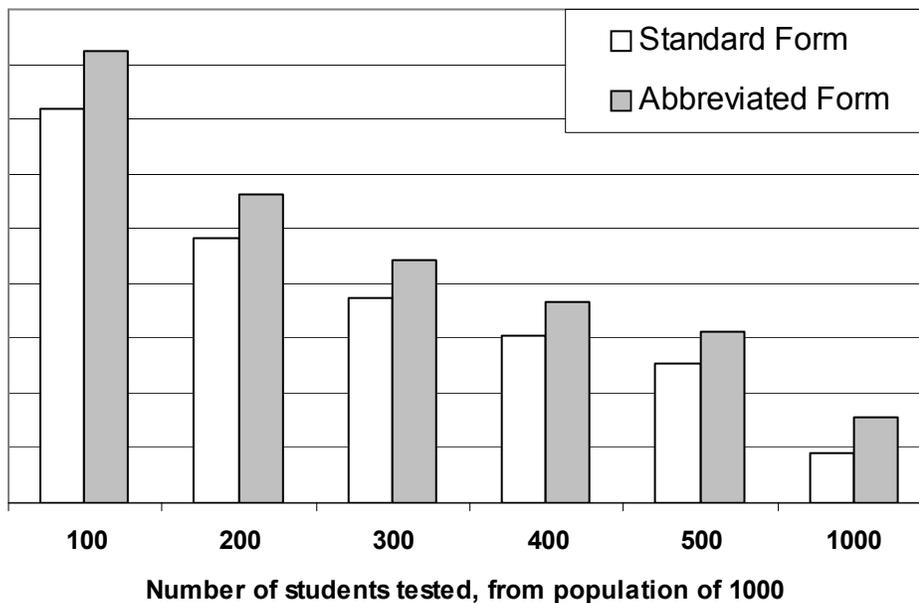
The “total population” is the group of students the test user would like the mean scores to generalize to. The distance between the confidence limits depends heavily on the number of students actually tested, but it also depends to some extent on the size of this population. Many test users are not interested in generalizing beyond the group of students who actually took the test. For those test users, the total population is simply the group of students tested. Their mean score will be very close to the true population mean (but not exactly the same, because the mean score might have been different if the specific questions on the test were different). The confidence limits will tend to be very close together. Other test users may want to generalize beyond the group of students tested. For those test users, the total population is larger than the group of students tested. The mean score of the students tested is more likely to differ appreciably from the true population mean, and the confidence limits will be farther apart.

When you request a Summary of Scaled Scores, the screen will display a box labeled, “Total Population.” The box will contain three buttons, for indicating the number of students in the population. Your choice will affect only the confidence limits on the report; it will not affect any other statistics. If you want the confidence limits to refer to a population that includes only the students actually tested, select “Generalize the report for this group of students.” If you want the confidence limits to refer to a very large group of students like those tested, select “Generalize report over a large group.” If you want to generalize to a specific group of students that is larger than the group actually tested, select “Generalize report for a specific number of students,” and enter the number of students in the box on the screen.

How Strongly Do the Confidence Limits Depend on the Number of Students Tested?

The following graph presents an example that shows how the distance between the confidence limits varies with the number of students tested. Notice that as the number of students tested gets larger, the distance between the confidence limits decreases. Also notice that the distance between the confidence limits is larger for the Abbreviated form than for the Standard form. A student taking the Abbreviated form answers only one-third as many questions as a student taking the Standard form. As a result, the students' scores on the Abbreviated form are less accurate, as estimates of the scores those students would earn if each student were tested with all possible sets of questions. Finally, notice that the distance between the confidence limits does not decrease to zero, even when all the students in the population are tested with the Standard form, because the group mean could be different (and probably would be slightly different) if the students were tested with a different edition of the Standard form.

**Relative Distance Between Confidence Limits
for Mean Scaled Score**



The reason the distance does not decrease to zero is that the confidence limits refer to the mean score that would result if all students in the population could be tested with all possible sets of questions. But, of course, each student answers only a fixed number of questions (e.g., 27 questions for a skills subscore on the Standard form). Therefore, each student's score is an imperfect measure of the score that student would earn if tested with all possible sets of questions. When the group mean is computed, these inaccuracies in the individual students' scores tend to average out, but not completely.

The ETS Proficiency Profile Essay

Institutions using the ETS Proficiency Profile test have the option to include an essay. The essay is administered only by computer. There is a 30-minute time limit for the essay itself, but the essay is preceded by a tutorial introduction that can take up to 15 minutes.

The ETS Proficiency Profile essay gauges students' ability to formulate and support a position on an issue and articulate that position clearly and effectively. Each student writes on only one topic. Each topic requires the student to express and support an opinion on a question with implications for the individual and for society as a whole. Some of the questions are of general interest; others are of particular interest to college students. Each question instructs the student to "support your position with specific reasons and examples from your own experience, observations and reading."

The students' essays are scored holistically on a six-point scale by *e-rater*[®], a computer program that analyzes written language. Before *e-rater* can begin scoring essays, it must be "calibrated," by having it analyze the language in a large group of student-written essays that have been scored by experts in the evaluation of student writing. Once *e-rater* is calibrated, it analyzes the language of each new essay and assigns it the score that the experts assigned to essays in which the language was similar.