

IDEAS for Assessment

For information on the IPT® and Language Assessment



INSIDE:

“Let the Numbers Do the Talking”

Get a sneak peak at the study that helped Arizona, Colorado, Maryland, New Mexico, Texas, and Utah determine scientifically based AMAOs for their ELLs.

(Story on page 4)

CONTENTS

Main article: Help is Here 1

IPT Q&A 2

From the Director of Assessment ... 2

Let the Numbers Do the Talking ... 4

IPT Testing Tip 5

Education News 5

New!

Poll: Heard Loud & Clear 5

HELP is Here

A study to determine “typical” yearly progress helps states determine benchmarks for their ELLs.

Accountability is a hallmark of the *No Child Left Behind Act* (NCLB). Schools must show that their students are making adequate annual progress in various subject areas. In the case of mainstream content classes, schools also are required to show that more of their students are reaching the “proficient level.” By the academic year 2013-14, 100% of the students must be grade-level proficient in reading and mathematics. There are more than four million English language learners (ELLs) in the school system, and NCLB requires that schools show that these students are making adequate progress in learning English, in addition to mainstream content. The federal government has not yet offered official guidelines on how this should be done.

Recognizing the responsibility this places on states, Ballard & Tighe conducted a study to help educators develop an accountability system for K-12 limited English proficient (LEP) students.

One State’s Example

Ballard & Tighe first examined how setting benchmarks would impact the number of students designated as making adequate progress in English language proficiency. This examination focused on one particular state

that used a single benchmark to show progress. Ballard & Tighe measured student results from the *IDEA Proficiency Test* (IPT), the state’s chosen testing system. Among the results, students at specific transition points (e.g., grade seven to eight) had a difficult time meeting the established benchmark. This pointed to developmental differences within the student population.

Researchers also found that individual students did not show uniform gains across the different areas of the IPT Oral, Reading, and Writing Tests. In other words, a student performed well in a certain area, but not quite as well in another. For this reason, the examination showed a relatively low percentage of students exhibiting *adequate progress* if they had to improve on all three tests according to the single benchmark set for each test. These results underscored that all students do not improve uniformly in oral, reading, and writing ability.

A Study to Help All States

Next, Ballard & Tighe researchers began a study of typical yearly progress that would benefit all educators using the IPT Tests. First, researchers retrieved and analyzed standardized proficiency scores on

(Continued on page 3)

PUBLISHED SEMIANNUALLY BY

Ballard & Tighe, Publishers
P.O. Box 219, Brea, CA 92822-0219
(800) 321-4332 www.ballard-tighe.com
e-mail: info@ballard-tighe.com

Dorothy L. Roberts
President/Owner

Roberta Stathis, Ph.D.
Vice President, Assessment &
Instructional Materials

Gary Buck, Ph.D.
Director of Assessment

Heera Kang
Editor

Kristin Belsher, Nina Chun, Patrice Gotsch,
Allison Mangrum, Sean O'Brien, Sherry Wicks
Editorial Staff

Jill Kinkade, Larry Newton, Katie Sprague
Marketing Staff

Liliana Cartelli
Art Director

Kathleen Styffe
Desktop Publishing Specialist

Cathy Sanchez
Printing Coordinator

Submissions for research articles, features, and
commentary are welcome; material may be
edited for length, style, and appropriateness.
All submissions must include the author's
name, address, and telephone number.

©2003 IDEAS for Assessment.
All rights reserved.

Reproductions in any form, in part or in
whole, are prohibited without written
permission from the publisher.

Ballard & Tighe is exhibiting at the following conferences:

Texas ABE
Oct. 28-Nov. 1
Ft. Worth, TX

Texas TESOL
November 7-8
El Paso, TX

New York TESOL
November 7-9
Rye, NY

La Cosecha
November 12-15
Albuquerque, NM

Carolina TESOL
November 13-15
Greenville, SC

**Northern New
England TESOL**
November 15
Manchester, NH

NABE 2004
February 5-7
Albuquerque, NM

Illinois TESOL
February 27-28
Chicago, IL

CABE 2004
March 4-6
San Jose, CA

TESOL 2004
March 25-29
Long Beach, CA

IPT[®] Q & A

Q: Can the same form of the IPT Oral Test be given more than once a year without compromising the results?

A: In general, repeated measures using the same form of any norm-referenced instrument should not take place within a minimum of four months between the first and second administrations. It is reasonable to test on a spring-to-spring or fall-to-spring testing schedule using the same instrument. Sufficient time has transpired over this period for the student to forget the specifics of the test form and to permit instruction to have an effect on the student's learning.

Q: Does the examiner have to speak the native language of the student being assessed?

A: No. Of course, it is always desirable before testing begins to put the student at ease. One way to do this is to converse with the student in the language in which he or she feels most comfortable. If the examiner does not have a command of the student's primary language, the examiner may need to take additional steps to put the student at ease. For example, the examiner might take extra time before testing begins to communicate with the student in English. It is imperative that the examiner be fluent in the language being tested.

E-mail your IPT questions to newsletter@ballard-tighe.com to be entered in a raffle for a free non-fiction historical reading book from the new *Explore the Ages* series. Type "IDEAS for Assessment" in the subject line.



From the Director of Assessment

Language assessment is a critical element in the *No Child Left Behind Act* (NCLB), and many districts and states have turned to Ballard & Tighe for help in developing accountability systems for English language learners (ELLs). In response, Ballard & Tighe's assessment team has conducted significant research to determine the typical yearly progress a student makes when taking the IPT language proficiency tests. You can read an overview of the findings of this study in this issue of *IDEAS for Assessment*. If you are interested in the full study, you can download a copy at www.ballard-tighe.com.

Our assessment team also has been working on a language proficiency test to meet all NCLB requirements. We are pleased to announce that the **IPT 2004 will be available for Spring 2004 testing**. Most of the current IPT test materials can be used to administer the test; only a few new components need to be purchased to use the NCLB-compliant IPT 2004.

In addition, we are continuing to develop the IPT 2005, formerly called the IPT+, which will be available for Spring 2005 testing. The IPT 2005, a revolutionary language testing system, can be aligned to individual state standards. IPT 2004 users will experience a seamless transition to the IPT 2005.

For more information about the latest assessment news at Ballard & Tighe, visit our new NCLB web site—nclb.ballard-tighe.com—where you will find current information about the IPT 2004 and IPT 2005. Ballard & Tighe is committed to making a difference in education!

Sincerely,

Gary Buck, Ph.D.

(Continued from page 1)

the IPT for students who have been tested annually for at least three years. Next, they used the data to set standards for annual progress and then explored a variety of ways to aggregate the individual data for school accountability.

PART 1: Determine Typical Yearly Progress

One goal of this study was to obtain empirically based estimates of “typical” student progress on the IPT Oral, Reading, and Writing Tests. To this end, Ballard & Tighe obtained student record data from districts in Arizona, Colorado, Maryland, New Mexico, Texas, and Utah. The data provided scores on the IPT Tests for consecutive grades in school. The data were subjected to the following quality control checks:

- The data needed to be recent (from within the last five years).
- Scores needed to be available for transitions of exactly one school grade (e.g., from first to second grade).
- The data were cleaned to remove cases with impossible score values.
- An outlier analysis was performed to identify extreme data values.

PART 2: Use Typical Student Progress to Set Benchmarks

The estimates of typical gains shown in this study provide an empirically determined basis for setting benchmarks for progress for the IPT Oral, Reading, and Writing Tests. (See “Let the Numbers Do the Talking” on page 4 for an excerpt from the study.)

The most striking conclusion, with respect to the mean gains, is that the mean gains appear to differ substantially, depending on the grade *transition point*. This is due to two factors:

1. Changes in the test forms. These are means of raw scores, so any transition point that corresponds to a change to a more difficult test will show a low or even negative “gain” as typical progress.

2. Actual developmental differences in progress.

Study Shows One Benchmark Does Not Fit All

The substantial difference in gains found in the study suggests that progress on the three tests is not uniform, so that more progress is achieved between some grade transitions and less between others. Consequently, a single benchmark for progress may be ill advised. A better solution is to have multiple benchmarks for each IPT Test, depending on the grade transition. If multiple benchmarks are eschewed in favor of a single one, then the result is likely to be a benchmark that is unrealistically high for some students while being unrealistically low for others.

Conclusions

This study has generated three general conclusions regarding the IPT and typical yearly gains: (1) This study clearly demonstrates the importance of obtaining empirically based estimates of typical gains for students; (2) The mean gains on the three IPT Tests used in this study differed substantially by transition point, suggesting that it may be prudent to have multiple benchmarks for each test, depending on the transition point; (3) This study also shows that discretion needs to be used when deciding on Annual Measurable Achievement Objectives (AMAOs). Specifically, it appears that while a given student may show a high amount of progress in one area of language, this does not necessarily coincide with a similarly high gain in other areas.

This study has been effective in establishing the school grade-level transition point as a factor in the gains students make in language development. It is certainly possible that there are other factors. One such factor may be the starting proficiency of the student. It is reasonable to expect that students of lower levels of English language proficiency may show more improvement after a year of instruction than those of a higher level of language proficiency. Eventually, all

Glossary

aggregate: combine

gains: change in scores—the score at year 2 minus the score at year 1

mean: the most often used measure of central tendency, due to its mathematical flexibility

outlier: extreme data value inconsistent with the model for the rest of the data

raw score: the direct score obtained immediately upon administration of the test

The most striking conclusion, with respect to the mean gains, is that the mean gains appear to differ substantially, depending on the transition point ... suggesting that it may be prudent to have multiple benchmarks for each test, depending on the grade transition point.

tests lose the ability to discriminate between students of very high levels of language proficiency—at some point, there is no room for improvement on the test. 🦋

For the full study, go to www.ballard-tighe.com. Ballard & Tighe can perform a similar study with data from your student population. Contact the educational consultant for your state or district at (800) 321-4332.

Authors

Gary Buck, Ph.D., Linguistics and Modern English Language. Dr. Buck has over 20 years of experience designing and developing language tests, with special expertise in tests of comprehension.

Sari Luoma, Ph.D., Applied Linguistics. Dr. Luoma has extensive experience in language testing, with special expertise in computer-assisted language testing and assessment of speaking.

Yeonsuk Cho, Ph.D., Educational Psychology. Dr. Cho has extensive background and experience in language testing, with a special emphasis on test development, rating scales for speaking and writing, and process-oriented writing assessment.

George Seretis, ABD, Measurement and Evaluation. Mr. Seretis has extensive background and experience in mathematical modeling and educational assessment with a special interest in categorical data analysis.



Let the NUMBERS Do the TALKING

Take a glimpse at the results of Ballard & Tighe's study on yearly progress. The following table summarizes the typical gains found for the IPT Reading Tests only. In addition to providing an overall gain estimate, the table below breaks down the gains by grade transition points. The full study analyzes the typical gains for the IPT Oral Tests and Writing Tests as well. (Go to www.ballard-tighe.com for the full study.)

Typical Gains for the IPT Reading Tests

The first column gives the test: IPT Early Literacy is used for grades K-1; IPT 1 for grades 2-3; IPT 2 for grades 4-6; and IPT 3 for grades 7-12.

The next column gives the grade transition, from one grade to the next.

The next column gives the number of students whose scores were looked at (i.e., the sample size); followed by the median gain (the gain made by the middle ranking student), and then the mean gain (the arithmetic average).

In the last column we provide confidence intervals. We want to know how confident we are that these results apply to other similar students, and this column gives the range of scores in which we would be 95% confident that other students would fall.

| IPT Test(s) | Transition Point | Sample Size | Median Gain | Mean Gain | 95% Confidence Interval for the Mean Gain |
|------------------------------------|------------------|-------------|-------------|-------------|---|
| Early Literacy | K→1 | 50 | 22 | 22.12 | 20.47 to 23.77 |
| Early Literacy→1 A & B | 1→2 | 51 | -27 | -24.98 | -27.60 to -22.36 |
| 1 A & B | 2→3 | 207 | 11 | 11.01 | 9.80 to 12.22 |
| 1 A & B→2 A & B | 3→4 | 183 | -3 | -2.34 | -3.48 to -1.20 |
| 2 A & B | 4→5 | 171 | 7 | 7.00 | 6.03 to 7.97 |
| | 5→6 | 131 | 6 | 5.85 | 4.57 to 7.12 |
| 2 A & B→3 A & B | 6→7 | 113 | 0 | -0.68 | -2.08 to 0.72 |
| 3 A & B | 7→8 | 91 | 4 | 5.69 | 4.29 to 7.09 |
| | 8→9 | 82 | 4 | 3.95 | 2.50 to 5.41 |
| | 9→10 | 74 | 4 | 4.46 | 2.90 to 6.02 |
| | 10→11 | 62 | 4 | 3.47 | 2.14 to 4.79 |
| | 11→12 | 53 | 3 | 2.66 | 1.41 to 3.91 |
| Overall | | 1268 | 4 | 4.02 | 3.42 to 4.62 |
| Overall (No Early Literacy) | | 1167 | 4 | 4.51 | 4.03 to 5.00 |

The gains on the Early Literacy test are quite high, 22 points, because at that young age children are making very rapid progress.

The transitions from one test to the next are all negative, which is to be expected given that the higher level test is always more difficult.

The gains between the IPT 1, IPT 2, and IPT 3 vary widely; ranging from 11 for grade 2 to 3 on the IPT 1 to 2.66 from Grade 11 to 12 on IPT 3. This suggests developmental differences.

Note: This table shows results expressed in raw score points. States represented are Arizona, Colorado, Maryland, New Mexico, Texas, and Utah.



IPT® Testing Tip

Ultimately, standardized tests can help educators determine an appropriate instructional plan for each individual student. In order to get accurate results, it's important to administer the tests correctly. A good testing system should be easy to administer and score, leaving little room for interpretation. IPT users often ask what to watch out for when administer-

ing the tests. The most serious mistake you can make in administering the IPT Tests is to give students credit for knowing things that they don't really know. Why? Because this can result in inaccurate placements and deny students services they need. We must be very careful not to give clues or extra help, or give students "the benefit of the doubt." In

essence, we must deny the urge to teach during standardized testing. If we change the procedure, reword a question, or accept an alternative answer, we may be invalidating the results. And, we want to have the most accurate results so that we have the appropriate information to make decisions that are in the best interests of our students!

EDUCATION NEWS

U.S. Department of Education Unveils New Web Site

The U.S. Department of Education has launched its new web site, which integrates *ED.gov* and *nclb.gov* into one comprehensive site. The new site also allows for easy access to NCLB news and information. To see the new site, visit: www.ed.gov/index.html

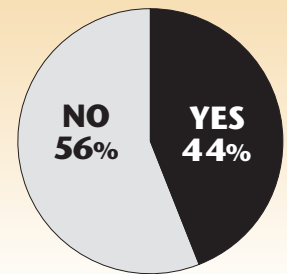
New Book Addresses Large-Scale Assessment of ELLs

As an outgrowth of concerns expressed by TESOL's task force on the *Elementary and Secondary Education Act (ESEA)* reauthorization, TESOL has published *Large-Scale Assessment of English Language Learners: Addressing Educational Accountability in K-12 Settings*, by Dr. Margo Gottlieb. The first section of this book provides working definitions of terms and reasons behind the renewed emphasis on testing and assessment. The second section aims to reshape the way in which ESL and bilingual education professionals measure the language proficiency of ELLs within a standards-based system. The third section outlines the parameters that constitute large-scale assessment. For more information, visit: www.tesol.org/pubs/offpress.html

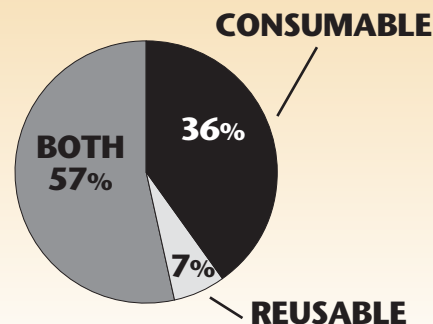
HEARD LOUD & CLEAR

This new section in *IDEAS for Assessment* features responses from educators around the country on questions related to the IPT Tests and English language assessment. It's a chance to see what other people think about issues important to you. So be sure to cast your vote. In this installment, we're highlighting responses from a nationwide survey that represented areas of AK, CO, MD, MI, MN, SC, TN, TX, and UT.

The IPT Tests are strictly paper and pencil tests, meaning there is no outside technology needed to administer or take the test. Would you like to have technology incorporated into the IPT?



Which type of test booklets does your school or district use?



NEXT POLL QUESTION:

Do the results of your English language proficiency tests help you with instructional planning?

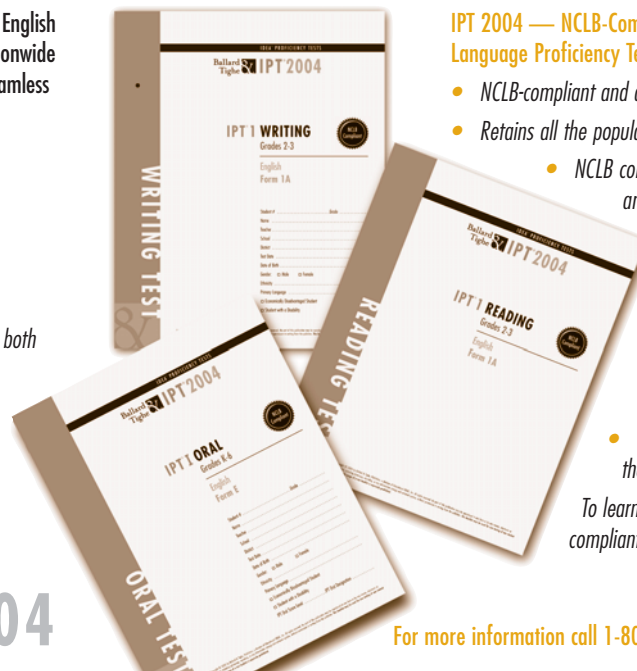
E-mail your response to newsletter@ballard-tighe.com to be entered in a raffle for a free nonfiction historical reading book from the new *Explore the Ages* series. Type "IDEAS for Assessment" in the subject line.

Well-respected language assessment series evolves to meet NCLB standards

Ballard & Tighe enhances its IPT Tests for limited English proficient (LEP) students to enable educators nationwide to navigate *No Child Left Behind* (NCLB) via a seamless migration path from today to 2005 and beyond.

IPT — Nationally Adopted Today

- Widely adopted by educators and assessment professionals
- Accurate and reliable language assessment
- Nationally normed tests measure proficiency in both English and Spanish
- Fully supported by Ballard & Tighe's long-valued, personalized customer service



IPT 2004 — NCLB-Compliant English Language Proficiency Tests for Spring 2004

- NCLB-compliant and available for Spring 2004 testing
- Retains all the popular testing and scoring features of the IPT
 - NCLB compliance is achieved through adapted scoring and reporting procedures
- Provides student scores in five key categories in English: Listening, Speaking, Reading, Writing, and Comprehension
- Will provide baseline data key to the development of Annual Measurable Achievement Objectives (AMAOs)
- IPT Manager 4: management software for the IPT testing system

To learn more about how the IPT 2004 is NCLB compliant go to: www.nclb.ballard-tighe.com