

August 2017 – Volume 21, Number 2

Testing as a Way to Monitor English as a Foreign Language Learning

Anthony Becker

Colorado State University, Fort Collins, USA
<tony.becker@colostate.edu>

Tatiana Nekrasova-Beker

Colorado State University, Fort Collins, USA
<t.nekrasova_beker@colostate.edu>

Tamara Petrashova

National Research Tomsk Polytechnic University, Tomsk, Russia
<petrashova@tpu.ru>

Abstract

This study was conducted at a large technical university in Russia, which offers English language courses to students majoring in nine different degree programs. Each degree program develops and delivers its own English language curriculum. While all degree programs followed the same curriculum development model to design language courses, each program incorporated a unique set of objectives pertaining to the subject matter of its discipline. The purpose of this study was to determine if progress tests could be a useful assessment tool to monitor the effectiveness of foreign language study throughout a University English Language Program (UELPP). Data from 600 English as a Foreign Language (EFL) students was analyzed using a repeated-measures ANOVA. The findings revealed that significant improvements in students' scores were gained throughout the first phase of the UELPP, which occurred over a two-year period. For the first time, the test data was used as a policy tool to introduce meaningful curricular adjustments, including revamping the instructional practices and methods of delivery to target a range of students' proficiency levels and establish the cut scores for a minimal level of language ability for Bachelor's degree students.

Introduction

According to some estimates, English is spoken by approximately 1.75 billion people worldwide (British Council, 2013). As Mufwene (2010) notes, much of the expansion of the English language can be attributed to the prescription of English as a second or foreign language in secondary schools of almost every country of the Outer and Expanding Circles today as to its usage as the primary lingua franca of business,

navigation, science and technology, and academia (p. 57). In Russia, it is estimated that only about five percent of the total population speaks English as a second language, which is in stark contrast to many other European countries (e.g., France, Germany, Netherlands, and Spain), where it is estimated that at least one-third of their respective population is bilingual or highly proficient in English (Abramova, Ananyina, & Shishmolina, 2013). Furthermore, in comparison to many other eastern European countries (including countries such as Bulgaria, Latvia, Poland, and Romania), Russian citizens tend to demonstrate far lower levels of English language proficiency (Education First, 2016). While there are many possible reasons to explain this circumstance (e.g., the geographical stature of Russia), many still argue that there is a strong need for Russia to further develop its English language programs, particularly at the university level (Abramova et al., 2013; Legasova, 2015).

Since the turn of the millennium, the Russian government has initiated several major efforts to improve the state of higher education and research in the country. In 2003, the Russian Higher Education System joined the Bologna Process. This process, which represents a series of agreements between European countries to ensure the comparability of standards and the quality of higher education qualifications (Reinalda & Kulesza, 2005), led to the appearance of more robust undergraduate and postgraduate degrees in Russia. Shortly thereafter (in 2006), the Russian government created a formal hierarchy of higher-education establishments, which led to the creation of a university ranking system for Russian universities, as well as the designation of special status for high-performing Russian universities (Smolensteva, 2015). More recently, in 2009, a system of universal examinations was introduced for all high school graduates (i.e., the Unified State Exam), whereby the results of these exams have become the sole basis for deciding university enrollment in Russia. While these three initiatives were not undertaken to address English language education in Russia only, it was during their creation that the importance of the English language for Russians was solidified.

As a result of these educational reforms, a number of Russian universities have developed action plans for establishing and promoting themselves as leading research institutions. Many of these action plans include goals and performance indicators for priority fields (e.g., business, computer science, engineering), as well as for English language education. For example, at Tomsk Polytechnic University (TPU), the administration has established several initiatives related to English, including the following goals: (a) improving the English language teaching system for TPU applicants, students, and staff; (b) introducing a documentation system in the English language; and (c) developing the university's bilingual social environment (with an emphasis on English) (TPU, 2013). The Russian Higher Education System believes that these efforts will help to close the perceived gap between Russia and its European counterparts.

Despite initiatives to promote English language education in Russia, the ability of universities to monitor students' progress in learning English has been somewhat hampered. Monitoring, which aims to ensure a constant supervision of a given process

so as to identify its correspondence to the desired result, can promote reflection on the results of educational and cognitive activities, as well as lead to possible corrections for the processes associated with them (Kaznachevskaya, 2013). It is often the case that efforts to monitor the effectiveness of English language programs in Russia are insufficient, as individual university departments are largely responsible for their own English language curricula (Tamara Petrashova, personal communication, March 15, 2017). Those who tackle the issue of monitoring student learning often refer to evidence of progress testing as an effective method (e.g., Bennett, Freeman, Coombes, Kay, & Ricketts, 2010; Schuwirth & van der Vleuten, 2012; van der Vleuten, Verwijnen, & Wijnen, 1996). The present study examines how progress testing was implemented to monitor student learning and improve instruction in an English language program at a Russian national research university.

Classroom-based language assessment

Language assessment, defined by Leung (2005) as the noticing and gathering of information about student language use in ordinary classroom activities, and the use of that information to make decisions about language teaching (p. 871), is a prominent component in most English language programs throughout the world. In second language classrooms, teachers implement assessments for many different reasons, including (but not limited to): (a) to monitor students' language learning; (b) to provide feedback to students; (c) to establish language-learning goals; and (d) to evaluate instructional effectiveness. Tests, as just one possible form of assessment, are most commonly used by teachers to serve the above-mentioned purposes (Miller, Linn, & Gronlund, 2012). While arguments can be made against their use (e.g., see Crowley, 2004; Gilbert, 2016; Popham, 1999), when effectively designed and implemented, tests can be a meaningful part of the assessment process, as they can help to enhance student learning and increase the effectiveness of teaching practices. This is particularly true of criterion-referenced (CR) tests, which, as Jamieson (2011) explains, "ha[ve] a well-established history as a means of focusing the attention of both teachers and learners on important areas of instruction" (p. 1).

Criterion-referenced testing

Since Glaser's (1963) coining of the terms *criterion-referenced* and *norm-referenced* in educational measurement, the prominence of CR tests has steadily grown, as they are seen as being more appropriate for answering questions about the actual achievement of students with respect to a particular domain (e.g., language learning). CR tests, in contrast to norm-referenced (NR) tests, which aim to compare an individual's performance against that of others, are intended to provide an evaluative description of the qualities which are to be assessed (e.g., an account of what pupils know and can do) without reference to the performance of others (Brown, 1988, p. 4). The purpose underlying CR tests is to determine whether an [examinee] can demonstrate specified real-world abilities. In this way, students are compelled to devote time and effort on the important aspects of a task and not to waste time on things they are not required to [know or] do (Johnstone, Patterson, & Rubenstein, 1998, p. 37).

While NR and CR tests do share some similarities (e.g., both can be used in instructional settings), there are a number of differences that help to distinguish these two types of assessment (see Clifford, 2016; Jamieson, 2011). For example, while NR tests are more commonly used to assess course-specific learning and to assign course grades, CR tests are more often used to assess mastery of specific learning outcomes, as well as curriculum-independent skills and higher-order, program-level instructional skills (Clifford, 2016, p. 225). In addition, NR tests usually result in the generation of a single, average (i.e., compensatory) score, while CR tests typically result in the generation of separate skill-specific (i.e., non-compensatory) scores. Furthermore, NR tests typically cover a large domain of learning tasks, whereas CR tests tend to focus more on a specified domain of learning tasks. Finally, as Clifford (2016) mentions, because of their independence from a curriculum, CR tests can be used to compare the abilities of students from different classes against a common set of external ability expectations (p. 225). For English language programs, many of which monitor their students' progress over the course of several semesters or years (Kaplan, 1997), CR assessments offer many distinct advantages for measuring progress not found in NR assessments.

Progress testing

As a form of CR assessment, progress tests are seen as being helpful in tracking students' improvement over time. Progress tests, which act as longitudinal feedback-oriented assessment tools (Schuwirth & van der Vleuten, 2012; van der Vleuten et al., 1996), are administered to the same cohort of students in the same program throughout their entire academic program of study. Additionally, they are usually administered at regular intervals (e.g., once per semester) and sample knowledge and skills expected of graduating students upon completion of their courses. Schuwirth and van der Vleuten (2012) argue that progress tests offer several advantages. Specifically, the authors report that they (pp. 26-28):

- are not restricted to a specific curriculum;
- reduce the examination stress experienced by students;
- complement traditional methods of assessment;
- positively influence the student learning process;
- are more predictive of future competence/performance; and
- add to the reliability of decisions.

Given the longitudinal and complementary nature of progress tests, their use also provides a unique snapshot of students' development throughout their course of study. Therefore, the information gleaned from progress tests serves to help make decisions about program advancement, instructional effectiveness and course design. Furthermore, progress tests can also be used formatively to help monitor an individual's growth throughout a period of instruction. In this way, the results of progress tests can be used to make decisions about feedback to students, remediation, and materials development. In either case, progress tests provide a wealth of information about individual learners, as well as about the program they are situated within.

While there is plenty of evidence to suggest that progress tests can be a useful addition to an existing assessment program, the research in support of their use has largely come from areas outside of language education and assessment, primarily within the fields of medicine and psychology (e.g., Bennett, et al., 2010; Dijksterhuis, Scheele, Schuwirth, Essed, & Nijhuis, 2009; Schaap, Schmidt, & Verkoeijen, 2011). To this point, research regarding the use of progress tests in language assessment has been limited, especially in comparison to the plethora of studies that have been conducted regarding other assessment and testing practices in the field. In addition, there is very little attention devoted to examining the English language assessment practices implemented in Russia. Given the perceived need for learning English as a second language in Russia (see Abramova, et al., 2013; Legasova, 2015), a greater awareness of the assessment practices implemented in English language programs at Russian universities is needed.

Present study

In light of the information presented above, the present paper attempts to investigate the English language assessment practices implemented in Russia. Specifically, this paper focuses on progress test data collected during the first of three stages (occurring from 2012-2014) of a required university English language program (UELP) offered to EFL learners studying at one of the tertiary institutions in Russia. The study sought to answer the following research question: *To what extent do EFL students demonstrate performance gains during the first phase of a UELP implemented at a prominent Russian university?*

Institutional profile and status of English

The university where the study was conducted is located in the southwest of Siberia, and is one of the leading polytechnic universities in Russia. The university consists of 7 scientific and educational institutes and offers four-year Bachelors degree programs and two-year Master's degree programs. The primary goal of these programs is to provide quality instruction to meet the educational needs of individuals, society and the State.

Since 1998, the university has emphasized English language teaching to ensure that future professionals are able to use the language to explore and adapt the best approaches and practices of their foreign peers, as well as to efficiently represent their own country in the foreign market. The university language departments and the faculty provide courses in the English language for students majoring in: Natural Science and Mathematics, Humanities, Applied Physics and Engineering, Electrophysics and Electronic Equipment, Economics and Management, Mechanical Engineering, Chemistry and Chemical Engineering, Thermal Power Engineering, Computer Science and Engineering. Since 2009, the UELP has been divided into three stages (see Table 1).

The first stage of the program focuses on developing mostly communicative and linguistic skills with a special emphasis on particular aspects of the language within the range of topics studied. The distinctive feature of the programs curriculum at the second

and third stages is an emphasis on learning English through specific academic content. Content-based instruction at these stages relies on the philosophy that learners acquire English by doing academic course work through the medium of that language (Pessoa et al., 2007). Students take specialized credit-bearing university courses in English and then prepare and defend part of their degree project in English. Thus, students acquire English by using it for both academic and professional purposes.

One of the distinctive features of the program's curriculum is the emphasis on teaching and learning English through the study of social as well as professional and academic content areas. With the goal of developing communicative language competence (*Common European Framework of Reference for Languages*, 2004), a great deal of emphasis is placed on teaching methods, especially those related to English for Specific Purposes (ESP). Furthermore, substantial emphasis has also been placed on learner-oriented instruction. As a result, students' needs have been given considerable attention in order to develop syllabi and ensure that a given course serves its target audience (Cowling, 2007).

Table 1. Structure of the University English Language Program

| Stage | Year of program | Degree | General goals |
|---------|-----------------|------------|--|
| Stage 1 | 1-2 | Bachelor's | To develop communicative competence in everyday situations relevant to home, university, social life, etc.; to prepare students for specialized ESP courses as well as to develop their academic English ability and study skills needed for success in undergraduate courses; to develop the English language proficiency needed to succeed in English-medium classrooms. |
| Stage 2 | 3-4 | Bachelor's | To develop communicative competence in the sphere of students' specialization and general areas of science. |
| Stage 3 | 5-6 | Master's | To develop discipline-specific competence as well as academic English ability and skills needed for success in career and science. |

In order to determine the relative value of language instruction methods employed by the UELP, and to monitor students' English language study, three progress tests were administered to all full-time students. The first progress test (PT1) was administered at the beginning of Stage 1, and was used to make entrance decisions for the UELP. The second progress test (PT2) was administered at the mid-point of Stage 1 (i.e., at the end of the first year of study), while the third progress test (PT3) was administered at the end of Stage 1 (i.e., at the end of the second year of study). Although the three tests did reflect some of the English language skills typically taught in the first two years of the

UEL, each test was intended to measure students' English language proficiency, not their language achievement, as they were not specifically linked to the English language curriculum implemented by the department.

Progress Test

Test standardization

According to Davies et al. (2002), a standardized test has to reflect a certain suite of characteristics, including: (a) rigorous development, trialing and revision, (b) standard procedures for administration and scoring of the test, (c) standard content in all test versions based on specifications, and (d) reliability of scores. Taken together, these characteristics are important for helping to ensure that a test is suitable for the purposes of comparability across large groups of test takers. As Davies et al. note, while all of these characteristics are of utmost importance for designing large-scale standardized tests, they should likewise be important considerations for any program that implements standardized tests. Therefore, the progress tests at the UELP were designed with these same characteristics in mind.

Specifically, the progress tests included authentic reading passages that closely resembled the types of English language texts that students were likely to encounter in their actual content-specific courses. They also included language skills (e.g., comprehension of main ideas and details, making inferences from context) and major content (e.g., academic vocabulary) targeted in the academic domain. Furthermore, the tests were piloted with a group of students recruited from the same population (n = 283), and test items were analyzed by computing item difficulty and item discrimination indices. Finally, internal consistency, which is an estimate of the reliability associated with how well the test items that reflect the same construct yield similar results (Bachman & Palmer, 2010), was determined for each of the tests using Cronbach's alpha (see below).

Test format

Each progress test included two sections: (1) *Reading* and (2) *Language Use*. Being restricted by practical constraints, such as time and cost, the test developers attempted to identify an essential core of language abilities that would be relevant to the range of academic situations in which students would find themselves. One of the main competencies that students were expected to acquire during the program was reading and comprehending general academic texts on technical topics as well as more specialized journals articles in their respective disciplines. Therefore, the *Reading* sub-test assessed the students' ability to understand written texts typical for the academic context. This test section was intended to tap such aspects of information processing as extraction of selected information, reading for the gist and for detailed information, and complex information processing including comprehension of implicit information. Another area of concern that was explicitly targeted in the language curriculum was students' ability to recognize and appropriately use morpho-syntactic constructions in an academic register (e.g., passive voice, nominalizations, the use of participles). Thus,

the *Language Use* sub-test assessed students' skills in operating morpho-syntactic constructions in a specific communicative context. Although the choice of tested skills related to the areas in which students need to succeed in an academic domain and which, therefore, were the most immediate needs of the program, we do acknowledge that the construct of the English language proficiency targeted in the progress tests was very narrow. Speaking was not included in the test, as it was assessed using other formative measures.

Overall, each progress test consisted of 50 multiple-choice items. The *Reading* section consisted of two short popular science or journalistic texts (250 to 350 words each), followed by five comprehension questions, for a total of 10 questions. The *Language Use* section included 30 fill-in-the-blank items which required students to complete gaps in given sentences with necessary grammar material, as well as 10 items requiring the identification of a mistake in one of the marked fragments of the sentence.

Each progress test was scored by assigning two points for each correct answer. Each wrong or absent answer was given a score of 0 points. No partial credit scores were assigned. The total score was the sum of scores for both sections, with a possible maximum of 100 points per test. See Table 2 for information about the number of questions in the sections, skills tested, and raw scores for each section.

Table 2. General Description of UELP Progress Tests

| Question Numbers | Skills tested | No. of questions | Question form | Max. score | % of total score |
|--------------------------------|---|------------------|--|------------|------------------|
| Section 1. Language Use | | | | 80 | 80% |
| 1-30 | Recognizing grammar to be correctly used in a given context | 30 | Multiple-choice, sentence completion | 60 | 60% |
| 31-40 | Recognizing grammar incorrectly used in a given context | 10 | Multiple-choice, error identification | 20 | 20% |
| Section 2. Reading | | | | 20 | 20% |
| 41-45 | Understanding main idea and/ or details | 5 | Multiple-choice, comprehension questions | 10 | 10% |
| 46-50 | Understanding vocabulary from context | 5 | Multiple-choice, word meaning identification | 10 | 10% |
| Total: | | 50 | | 100 | 100% |

Students had a total of approximately 256 class hours (4 hours/week) in the UELP, and the time periods between the three test administrations were roughly equal. Thus, it meant that after PT1 the students received English instruction for about 128 hours before taking PT2, and PT3 occurred after about the same number of hours following the completion of PT2.

Prior to administering the tests, a standard-setting study to establish the level of the language and skills tested was carried out by UELP teachers and test administrators, each with more than 10 years of experience in English-language instruction. According to the panelists' judgment, the language used and skills tested in the questions corresponded to the levels of foreign language proficiency specified by the *Common European Framework of Reference for Languages* (Council of Europe, 2014).

Each administration of the progress test was carried out on computers using the Moodle Course Management System. The testing took place in a classroom equipped with personal computers with Internet access. All test-takers were capable computer users since they had already studied the basics of computer science. While the paper-based versions of the test were also available, none of the participants included in the study took the paper and pencil test. The testing procedures were monitored by proctors who received special training to administer the tests.

The testing session lasted 45 minutes, not including time for instructions. The use of dictionaries, other study and reference materials, mobile communication devices and other sources of information during the testing time was not allowed. While performing the test, students could take notes on blank paper provided by the proctors. However, students' notes were not taken into account in the scoring procedures. Scoring was performed automatically for each progress test. Students' total score was reported immediately following the test, and was registered and stored by the online system. Students were given only one attempt to take the test.

In order to build an individual progress report for individuals studying within the program, students' scores for the three progress tests were compared. Specifically, scores for PT1 were compared to scores for PT2, and those scores were then compared to the scores for PT3. Having scores from three separate test administrations made it possible to measure students' progress throughout the program and to assess students' end-of-program English-language proficiency with respect to the skills tested.

Data Analysis

There were three test administrations for students who entered the university in 2012. PT1 was administered in September 2012, PT2 in May 2013, and PT3 in March 2014. For these three tests, one and the same test battery (compiled out of 500 bank items that were identified by their content specifications and for which item statistics were available) was administered. The number of students who took part in the UELP testing was as follows: PT1 -1813 (87%); 2- 1547 (79%); 3 -1477 (83%). However, the scores

of only those students ($n = 1154$) who participated in all three test administrations were considered for the analysis.

In order to determine the reliability of the scores for each test, internal consistency was calculated for each of the test forms using Cronbach's Alpha. According to Kline (2000), alpha values ranging from 0.7 to 0.9 are adequate, while values at, or above, 0.9 are desirable for high-stakes testing. As the UELP progress tests were considered to be relatively low-stakes testing, reliability coefficients at, or above, 0.7 were considered adequate.

In addition, a repeated-measures analysis of co-variance (RM-ANCOVA) test was used to compare the mean test scores of 600 randomly selected examinees across the three test administrations. Test administration served as the within-subjects variable and the mean scores for the three test administrations served as the dependent variable. ANCOVA is particularly useful in situations when the dependent variable could be adjusted for differences in the covariate(s) (Mayer, 2013). For the present study, the group mean scores were adjusted to account for the different test forms that were administered (i.e., the covariate).

Results

Using Cronbach's Alpha, internal consistency was first calculated. Overall, the reliability coefficients across the different test forms ranged from .796 to .893, which were considered to be adequate for these progress tests. In addition, the mean scores of 600 examinees were compared across the three test administrations. The descriptive statistics are presented in Table 3.

In order to ensure that the RM- ANCOVA test was being used appropriately, certain assumptions had to be met (see Tabachnick & Fidell, 2007). The kurtosis and skewness values ranged between +/- 2, suggesting that the assumption of normality had been satisfied. Furthermore, the result of Mauchly's Test of Sphericity was non-significant [$X^2(2) = 4.81, p = .203$], indicating that there was equal variance across all three administrations. As all assumptions were met, the use of RM-ANCOVA was deemed appropriate for the present study.

Table 3. Descriptive Statistics for PT1, PT2, and PT3 ($n = 600$)

| Test | Mean (for group) | SD | Min. – Max. |
|---------------|------------------|-------|---------------|
| PT1 | 37.11 | 13.73 | 10.00 – 84.00 |
| PT2 | 41.61 | 15.53 | 14.00 – 88.00 |
| PT3 | 47.50 | 17.92 | 12.00 – 96.00 |
| Total average | 42.07 | 16.37 | — — |

The RM-ANCOVA test revealed a statistically significant main effect, $F(2, 597) = 54.27, p < .05$, indicating that the mean total scores for the tests were not the same for all three test administrations. Post-hoc comparisons, using Tukey HSD procedures, were used to determine which pairs of the three group means differed. As Table 4 shows, the scores from the first test administration (PT1) were significantly lower than the scores from the second and third test administrations (PT2 and PT3). The effect sizes for these significant pairwise differences were 2.35 and 4.87, respectively. In addition, the mean score difference between PT2 and PT3 was found to be significant. The effect size for this significant difference was 2.55.

Table 4. Tukey HSD Post-Hoc Results for Three Test Administrations

| Mean Differences ($\bar{X}_i - \bar{X}_k$) (Effect Size is indicated in parentheses) | | | | |
|---|-------|------------------|-----------------|----|
| Test | Mean | 1. | 2. | 3. |
| 1. PT1 | 37.11 | — | | |
| 2. PT2 | 41.61 | 4.54* (2.35) | — | |
| 3. PT3 | 47.50 | 10.96* (4.87) | 5.89* (2.55) | — |

* $p < .01$

Discussion

The present study focused on progress testing as one possibility to monitor the development of students' language abilities. The results revealed that students' mean test scores improved significantly from one progress test to the next (i.e., PT1 PT 2 PT3) over a period of four academic semesters. These findings partially support similar research (e.g., Elder & O'Loughlin, 2002) that also investigated the relationship between English language study and score gains on a standardized test. There are several possible explanations for the findings in the present study.

As a whole, students who were in the first stage of the UELP appeared to experience greater gains than students in the second stage of the program. Elder and O'Loughlin (2002) explain that this is likely because the proficiency that one starts with is the most constant indicator of how far one is likely to travel over the course of their language studies (p. 226). In other words, those students who began the current study with a lower level of English language proficiency had a higher ceiling for growth over the course of the study compared to students who began the study with a higher level of proficiency.

Furthermore, the large gains in students' overall scores could also be explained by a variety of other factors outside of the curriculum. For instance, a considerable number

of students in the present study sought additional assistance (e.g., tutoring) on top of their English language instruction. In addition, not surprisingly, those students who regularly attended their English language classes performed far better on the second and third progress tests. Future research could build on the present study by considering other factors outside of the curriculum that are likely to influence test score gains, such as learning experiences with peers, parental support, educational background, and motivation (Elder & O'Loughlin, 2002; Shavelson, et al., 2010).

Since the main purpose of the study was to examine if the results of progress tests could provide additional insights into the quality of English language courses offered at the university, the findings of the study had direct implications for classroom instruction, curriculum development, and policy making.

On a classroom level, test results were used to inform language instruction, including adjusting instructional practices and methods of delivery to target a range of proficiency levels that are often present in a given language classroom. As several instructors reported during interviews, using test scores to group students during classroom activities had been a useful strategy to ensure that the needs of all students were met. Also, since language instructors across university language programs were provided access to the overall summary of the results describing the performance of their students in relation to other departments and specializations, this information was used to identify the specific linguistic structures and sub-skills that appeared to be challenging for each particular group of students. Once the salient points were identified, the instructors then sequenced the material in terms of the difficulty level and dedicated additional classroom time to address those points. Depending on the specific needs of the students, the type of instructional support differed in each class and covered a range of activities from contextualized presentation of the target material to guided practice to providing opportunities for more creative use of the language and fluency development. Finally, multiple feedback sessions have been conducted with language instructors from various English programs at the university in attempts to discuss the goals of the progress testing system, its place in the overall educational process, and how the results should be interpreted and what types of decisions can be made based on those results. Following these discussions, the structure of the progress testing system has been revised as well to include an additional section on Listening, an important sub-skill of the functional language ability that is targeted during language instruction across all university English language programs.

In addition to direct implications for classroom instruction, the study provided justification for the use of progress tests as one of the ways to monitor students' language development in different university language programs and to provide remediation for students. Students' performance on each of the three tests was analyzed and the scores were reported back to the students, along with qualitative feedback and recommended instructional modules that were developed to provide additional language instruction (including explanations, examples, and pedagogical activities) on the most challenging content targeted in the progress tests. Currently, 13 modules have

been developed focusing on 10 language usage aspects (e.g., the use of passive voice in an academic register) and three aspects of reading comprehension (e.g., understanding main ideas), all targeting B1-B2 levels on CEFR (for modules specification, see Petrashova & Yagovkina, 2013). All modules have been designed for on-line delivery to motivate students to work independently outside of the classroom. The content targeted in the modules has been identified based on the item analysis of test items, including item difficulty, which was conducted for the entire pool of test-takers. Since the proficiency levels of the students vary, the decision was made to provide all explanations included in the modules in Russian to ensure that the content was accessible to all students, regardless of their proficiency level in English. All developed modules are hosted on a web-based course support system and are open to all individuals who have taken the progress test. Once the results of the progress tests and the feedback associated with the scores become available, students receive unlimited access to the modules, so that they are able to go through the material at their own pace from any location, as long as they are logged into the university system.

Finally, in terms of more global implications, the results of the tests were used to introduce several policy-making decisions at the university. One such decision was to incorporate the results of the progress tests in the annual evaluation of language departments performed by the Vice-Rector for Student Affairs, which offers additional evidence about the quality of language instruction provided by the department. In addition, the discussion of the test results at the university council has resulted in the decision to accept the cut-point of the proficiency test (of 50 and above) administered at the end of the second year (PT3) of study as a minimal level for Bachelor's degree students, as well as an admission requirement to pursue master degree programs at the university. Finally, the results of the tests are also reviewed by individual departments to pre-screen students for participation in international exchange programs and research activities that require a certain level of English proficiency.

Limitations

The results of this study should be interpreted with caution for several reasons. First, while standardized tests can be used to chart students' language development, it is important to remember that language growth patterns should never be based solely on test scores. Instead, language programs should also incorporate informal assessment methods to monitor progress and determine whether or not students' language skills are improving (Short, 1993). Additional insights can be gained from evaluating students' performance in the classroom concurrently with their performance on progress tests. Furthermore, information about students' performance can be gleaned from instructors, as well as from students themselves. Such information, along with performance on progress tests, would likely be more revealing for informing instruction and designing remedial materials for learners (Lee & Sawaki, 2009).

Second, the progress tests included in the study focused on a rather limited set of skills that, no doubt, imposed an important limitation on the evaluation of test takers

communicative competence. Because the tests were designed with some practicality constraints (e.g., time availability, programming, resources), it includes only multiple-choice items that provide a certain ease of recognition and guessing success. Therefore, determining the relationship of test performance largely based on multiple-choice items to any “real-world” criterion requires further empirical examination to determine the extent of its correlation with other measures.

Conclusion

The present study explored if the use of progress tests could provide additional information about the quality of English language instruction at a large public university in Russia. The results indicated that, overall, students showed significant increases in test scores across the three test administrations, which, in turn, led to a number of important administrative decisions made at the university. At the same time, the results of the study also highlighted the need for broadening the construct of the English language proficiency by incorporating productive language skills.

About the Authors

Anthony Becker is an Assistant Professor in the English Department at Colorado State University. He has been teaching in the TEFL/TESL program there since 2012. He holds a PhD in Applied Linguistics from Northern Arizona University. Aside from work dealing with second language assessment, his other research and teaching interests include language for specific purposes, second language writing, and computer applications in applied linguistics. When not working, he enjoys an occasional hike or run in and around Fort Collins, Colorado.

Tatiana Nekrasova-Beker is an Assistant Professor in Applied Linguistics and TEFL/TESL at Colorado State University where she is currently teaching graduate courses in the TEFL/TESL program, including Teaching English as a Foreign/Second Language, Theories of Foreign/Second Language Learning, and Curriculum Development in English for Specific Purposes. She holds a doctorate in Applied Linguistics from Northern Arizona University. Her research interests include usage-based approaches to L2 acquisition, the role of formulaic language in fluency and syntactic development, project-based methods in L2 instruction, and corpus-based analyses of ESP texts.

Tamara Petrashova heads the Department of Quality Assurance in Foreign Language Training at TPU. She received her PhD in Philology from Ivanovo State University (Russia) in 2006. Dr. Petrashova’s current research interests cover the influence of assessment on learning teaching process, ESP assessment, advanced methods in EFL, project-based learning and teaching, and various topics in applied linguistics research, including vocabulary acquisition, methods and techniques in corpus linguistics.

References

- Abramova, I., Ananyina, A., & Shishmolina, E. (2013). Challenges in teaching Russian students to speak English. *American Journal of Educational Research*, 1(3), 99-103.
- Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice*. Cambridge: Cambridge University Press.
- Bennett, J., Freeman, A., Coombes, L., Kay, L., & Ricketts, C. (2010). Adaptation of medical progress testing to a dental setting. *Medical Teacher*, 32, 500-502.
- British Council (2013). *The English effect: The impact of English, what its worth to the UK and why it matters to the world*. Retrieved from <https://www.britishcouncil.org/sites/default/files/english-effect-report-v2.pdf>
- Brown, S. (1988). Criterion-referenced assessment: What role for research? *British Journal of Educational Psychology*, 3, 1-14.
- Clifford, R. (2016). A rationale for criterion-referenced proficiency testing. *Foreign Language Annals*, 49, 224-234.
- Common European framework of reference for languages: Learning, teaching, assessment* (2004). Council of Europe: Cambridge University Press. Accessed on September 2014, from http://www.coe.int/t/dg4/linguistic/Source/Framework_en.pdf.
- Cowling, J. D. (2007). Needs analysis: Planning a syllabus for a series of intensive workplace courses at a leading Japanese company. *English for Specific Purposes*, 26, 426-442.
- Crowley, C.J. (2004). The ethics of assessment with culturally and linguistically diverse populations. *The ASHA Leader*, 9, 6-7.
- Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T., McNamara, T. (2002). *Dictionary of language testing*. Cambridge: Cambridge University Press.
- Dijksterhuis, M. G. K., Scheele, F., Schuwirth, L. W. T., Essed, G. G. M., & Nijhuis, J. G. (2009). Progress testing in postgraduate medical education. *Medical Teacher*, 31, 464-468.
- Elder, C., & O'Loughlin, K. (2002). *Investigating the relationship between intensive English language study and band score gain on IELTS*. Unpublished report (Volume 4, Report 6).
- Frey, Bruce B., Schmitt, Vicki L., & Justin P. Allen (2012). Defining authentic classroom assessment. *Practical Assessment, Research & Evaluation*, 17(2). Accessed October 2014, <http://pareonline.net/getvn.asp?v=17&n=2>
- Gilbert, E. (2016). *Why assessment is a waste of time*. Retrieved from <https://www.insidehighered.com/views/2016/11/21/how-assessment-falls-significantly-short-valid-research-essay>
- Glaser, R. (1963). Instructional technology and the measurement of learning outcomes. *American Psychologist*, 18, 519-522.

- Jamieson, J. (2011). Assessment of classroom language learning. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 768-785). New York, NY: Routledge.
- Johnstone, R., Patterson, J., & Rubenstein, K. (1998). *Improving criteria and feedback in student assessment in law*. Sydney: Cavendish Publishing.
- Kaplan, R.B. (1997). An IEP is a many-splendored thing. In M.A. Christison & F.L. Stoller (Eds.), *A handbook for language program administrators* (pp. 3-19). Provo, UT: Alta Book Center.
- Kaznachevskaya, L.V. (2013). Didactic monitoring in the process of teaching the English language. *Modern Research of Social Problems*, 7, 1-14.
- Kline, P. (2000). *The handbook of psychological testing* (2nd ed.). New York, NY: Routledge.
- Lee, Y.W., & Sawaki, Y. (2009). Cognitive diagnostic approaches to language assessment: An overview. *Language Assessment Quarterly*, 6, 172-189.
- Legasova, T.A. (2015). English in Russia: To learn or not to learn? That is the question. *Asian Social Science*, 11, 231-234.
- Leung, C. (2005). Classroom teacher assessment of second language development: Construct as practice. In E. Hinkel (Ser. Ed.), T. McNamara, A. Brown, L. Grove, K. Hill, and N. Iwashita (Eds.), *Handbook of research in second language teaching and learning* (pp. 869-888). Mahwah, NJ: Lawrence Erlbaum Associates.
- Mayer, A. (2013). *Introduction to statistics and SPSS in psychology*. New York, NY: Pearson.
- Miller, D. M., Linn, R. L., & Gronlund, N. E. (2012). *Measurement and assessment in teaching* (11th ed.). New York, NY: Pearson.
- Mufwene, S.S. (2010). Globalization and the spread of English: What does it mean to be Anglophone? *English Today*, 26(1), 57-59.
- Pessoa, S., Mellon, C., Hendry, H., Donato R., Tucker, G. R., Mellon C., & Lee, H. (2007). Content-Based Instruction in the Foreign Language Classroom: A Discourse Perspective. *Foreign Language Annals*, 40(1), 102 -121.
- Petrashova, T., & Yagovkina, M. (2013). *Module specification for the English progress test*. National Research Tomsk Polytechnic University, Tomsk.
- Popham, J.W. (1999). Why standardized tests don't measure educational quality. *Educational Leadership*, 56, 8-15.
- Reinalda, B., & Kulesza, E. (2005). *The Bologna process: Harmonizing Europe's higher education* (1st ed.). Barbara Budrich Publications.
- Schaap, L., Schmidt, H., & Verkoeijen, P. J. L. (2011). Assessing knowledge growth in a psychology curriculum: which students improve most? *Assessment & Evaluation in Higher Education*, 1-13.

Schuwirth, L.W.T., & van der Vleuten, C.P.M. (2012). The use of progress testing. *Perspectives on Medical Education*, 1, 24-30.

Shavelson, R.J., Linn, R.L., Baker, E.L., Ladd, H.F., Darling-Hammond, L., Shepard, L.A., Barton, P.E., Haertel, E., Ravitch, D., & Rothstein, R. (2010). *Problems with the use of student test scores to evaluate teachers*. Retrieved on May 15, 2015 from <http://www.epi.org/publication/bp278/>.

Short, D. (1993). Assessing integrated language and content instruction. *TESOL Quarterly*, 4, 627-656.

Smolentseva, A. (2015). Russian system of higher education and its stakeholders: Ten years on the way to congruence. *Higher Education Dynamics*, 44, 215-236.

Tabachnick, B.G., & Fidell, L.S. (2007). *Using multivariate statistics*. Boston, MA: Pearson, Allyn & Bacon.

Tomsk Polytechnic University (2013). *Action plan on the implementation of university programme for promoting the competitiveness among world's leading research and educational centers*. Tomsk, Russia: Tomsk Polytechnic University Press.

van der Vleuten, C. P. M., Verwijnen, G. M., & Wijnen, W. H. F. W. (1996). Fifteen years of experience with progress testing in a problem-based learning curriculum. *Medical Teacher*, 18(2), 103-109.

© Copyright rests with authors. Please cite *TESL-EJ* appropriately.