EnglishUSA Journal

Volume 7 Article 6

2022

Measuring Student Achievment in an Intensive English Program

Seo Hyun Park
Divine World College, spark@dwci.edu

Measuring Student Achievment in an Intensive English Program

Measuring Student Achievement in an Intensive English Program

by Seo Hyun Park

Introduction

This paper introduces a case of statistical analysis that a college-governed intensive English program (IEP) did to assess student development as part of its program evaluation. Rubrics and final exams are a typical data source for research on student progress in IEPs (Juffs, 2020). This approach, however, may not show a program-wide student achievement trend over time, especially in IEPs where rubrics, textbooks, and final exams are not standardized and unified across levels and courses but rather dependent on individual instructors' discretion. Researchers suggest multiple methodologies of data collection and rich contextualization of data analysis as a good practice of language program evaluation (Hamdoun, 2021; Lynch, 1996; Shawer, 2013). In the IEP at Divine Word College (DWC) in Epworth, Iowa, structured evidence was needed to understand a student achievement trend besides a routine review of unstructured, qualitative data available from student opinion surveys, exit interviews, and comments from student course evaluations[1]. This missing puzzle piece in the IEP at DWC for an extended period of time seems partially because IEP faculty and administrators are extremely busy and thus have little time to sit and learn and do statistics; another reason may be related to statistics anxiety generally found among language educators (Brown, 2013).

Nevertheless, even taking a simple look at placement-exit score differences may lead an IEP off to a good start to triangulate data resources, confirm its experiential and qualitative findings, and establish a healthy program evaluation cycle of context, input, process, and product (Stufflebeam, 2003). The purpose of this paper is to share a practical example of student achievement recording and reporting practice with IEP stakeholders, agreeing with Norris' (2009) observation as the following:

> A commonly cited gap in program evaluation work is the lack of published reports on how evaluation happens, largely due to the fact that evaluation reports tend to be produced for evaluation clients rather than a broader academic public. (p. 9)

Program Background and Research Questions

The DWC IEP revamped its program format in 2018, changing the number of proficiency levels from five to eight and the length of each level from a semester to a half semester. Three years later, in 2021, was when experience, data, and faculty discussion were gathered enough for preliminary research on how students were doing in the newly formatted program. The purpose of this research, guided by program evaluation researchers (e.g. Scriven, 1967), was twofold: (1) summative reporting purposes to college administrators and eventually to the language program accreditor (CEA, or the Commission on English Language Program Accreditation) and (2) formative improvement purposes within the IEP.

[1] See Powers (2020) for discussions of structured vs. unstructured data strategy.

For both purposes, two questions needed to be addressed:

1. Do students in the new half-semester program achieve English language skills better than those in the old semester-based program?

2. Which English language skill of students' reaches a college-entry proficiency level the fastest and the slowest in the new half-semester program?

Answering these questions was an important step for the DWC IEP's continued program development, planning, and review. Data-driven implications were much needed to answer other subsequent curricular or administrative questions such as how long it actually takes for Level 1 students to exit the program, how many instructional hours are to be allotted for each language skill, which standardized tests assess student learning most accurately yet conveniently, and which textbooks are to be used, for how many years, and with what supplementary materials, to name a few.

To answer these questions, survey and interview data were not considered since they mostly shed light on student perception, not student achievement. Course grade and in-class rubrics analyses were not the best option, either, given that DWC IEP faculty develop and use their own rubrics and flexibly change assessment tools and percentages by course and level as well as by types of students they serve in a particular term. Instead, placement and end-of-term test scores were analyzed for the present study given the fact that, in the DWC IEP, three tests are used both for the initial placement and for the promotion to the next level at the end of every academic term: in-house English writing test (EWT), in-house English Speaking Test (EST), and the Cambridge Michigan Language Assessment English Placement Test (CaMLA EPT).

The DWC IEP is aware of strengths and limitations of using locally developed tests (EWT and EST) and direct assessment techniques (published tests such as CaMLA) for placement and promotion, and further, for research on student achievement. Juffs' (2020) study, for instance, finds out that "the writing test is not terribly useful in placing learners in appropriate classes based on intended student learning outcomes" (p. 76). California DOE's (1994) Student Oral Language Observation Matrix, from which DWC IEP's EST rubrics were adapted, is originally for placement only, not for progress assessment. So is the CaMLA EPT, and the Michigan Language Assessment does promote other test products that specifically assess student progress and for high-stake decisions (i.e., MTELP or MET[2]). Despite these caveats, the DWC IEP intentionally uses the same test format for both existing and new students so that their language proficiency is measured fairly and reliably across levels in each term. Allen (2004) argues that practicality cannot be ignored when developing methods to assess

[2] The MTELP (Michigan Test of English Language Proficiency) is a progress and placement test which is sold directly to institutions and administered at institutions without external monitors. The MET (Michigan English Test) is administered only at authorized test centers and used for certification purposes.

student achievement. As long as end-of-term test scores are not used as a single factor of promotion decisions, using the same standardized tests for placement and promotion is a practical way to measure student achievement.

Methods

Student scores of EWT, EST, and four sub-scores of CaMLA EPT (listening comprehension, grammar, vocabulary, and reading) were collected from December 2017 to March 2021 and sorted out on an Excel spreadsheet. EWT scores from three different scoring systems and rubrics over time were equalized by entering the achievement rate (%) of each student score out of a full score of 5.0 in 2017, 25.0 in 2018, and 9.0 since 2019. Since the data set is small (28 students who began English as a second language (ESL) study either in August 2017 or in January 2018 and 70 students who began ESL study in August 2018 or later), independent sample t-tests were conducted with a p-value of 0.05 to assure the validity and reliability of data analysis.[3] A test of correlation was conducted with the same p-value to examine the progress of and relationships among six language skills (writing, speaking, listening, grammar, vocabulary, and reading). Stata 17 was used for these tests.

Data Analysis and Discussion

Independent sample t-tests show that students in the half-semester program achieve higher writing, speaking, vocabulary, and reading proficiencies for a four-month period than students in the semester program. On the other hand, students in the half-semester program do not achieve listening and grammar proficiencies as high as students in the semester-long program. Table 1 presents absolute score differences between two sets of English tests that were taken with a four-month interval after a semester or two eight-week terms of ESL education. Figure 1 converts numbers in Table 1 to the percentage out of the full score of each test for a visually convenient comparison.

	w1	w2	s1	s2	l1	l2	g1	g2	v1	v2	r1
1	3.722222	4.277778	13.38889	15.72222	11.33333	15.33333	10.33333	13.94444	8.388889	9.833333	4
	3.5	4	14	15.5	11	16	11	14	8	11	3.5
	1.526455	1.190924	3.070522	2.191039	4.014679	3.531372	3.162278	2.071437	2.428722	3.053445	2.249183
2	3.986667	5.46	14.88667	17.83333	12.34667	14.93333	11.33333	13.32	9.693333	11.42667	5.56
	4	5.5	16	18	12	15	12	14	10	12	5
	1.982991	1.857709	4.452887	3.934715	4.746246	4.54507	4.214946	3.38981	3.759576	3.966663	2.964839
1	3.935484	5.231183	14.59677	17.42473	12.15054	15.01075	11.13978	13.44086	9.44086	11.11828	5.258065
	4	5	14	17	12	16	12	14	9	11	5
	1.898549	1.805124	4.247927	3.747424	4.610824	4.352647	4.036764	3.177579	3.567554	3.84439	2.896412

Table 1. Average score changes for a four-month period in semester program and half-semester program[4]

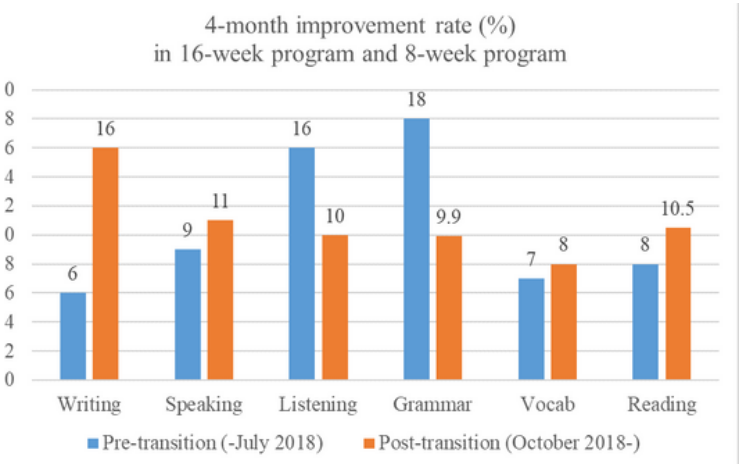


Figure 1. Improvement rate in pre-transition and post-transition periods

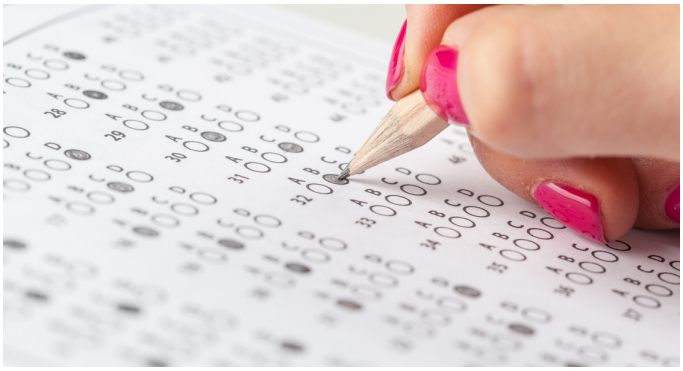
[3] A simple descriptive analysis (e.g. comparing means) is meaningful only with a large dataset such as those from Data.gov, HealthData.gov, Harvard Dataverse, World Bank Microdata Catalog. For a small data set, statistical tests are used to determine the probability and representativeness of the data.

[4] Row 1 = mean; Row 2 = median; Row 3 = standard deviation; Transition 1 = 16-week program, Transition 2 = 8-week program; w1, s1, l1, g1, v1, r1 = writing, speaking, listening, grammar, vocabulary, and reading scores obtained in the beginning of the analyzed four-month period; w2, s2, l2, g2, v2, r2 = writing, speaking, listening, grammar, vocabulary, and reading scores obtained four months later from w1, s1, l1, g1, v1, and r1

Factors worth discussing regarding this progress difference between old and new IEP structures include student learning outcomes and textbooks that are finely aligned with the Common European Framework of Reference for Languages, changes in weekly instructional hours (oral communication from 5 to 10, reading from 5 to 3), and more opportunities for students to take EWT, EST and CaMLA in the new half-semester program.

This interpretation has a limitation, however, the score differences only with a four-month interval were compared due to the limited data available from the old semester-long program. For instance, the EST began to be offered in the IEP at DWC only in January 2018; the Michigan Language Assessment began to offer CaMLA EPT subscores on its online portal in December 2017. With this taken into consideration, a reasonable interpretation of Table 1 and Figure 1 will be that both old and new programs did help students learn English but in varied degrees across language skills. With this finding, the program reorganization at DWC in 2018 is justified.

When only the new-program score data are analyzed, a longer-term view is available. Figure 2 shows an increasing pattern of student scores over terms in the half-semester (i.e., eight-week) program.



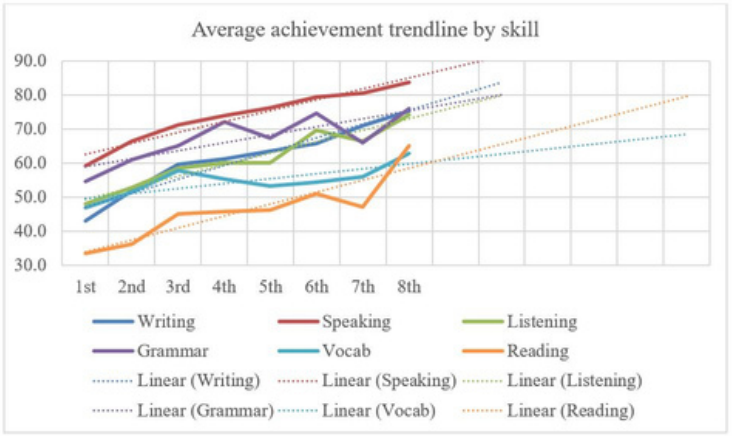


Figure 2. Average achievement trendline by skill over terms in the eight-week program

Looking at the starting point of each solid line in Figure 2, it is noted that most students join the DWC IEP with some knowledge of grammar and speaking proficiency. Nevertheless, they are not the easiest skills to improve—their slope values from the equation are 2.37 and 3.21, respectively, which are lower than those of other skills except reading.[5] Reading and writing, on the other hand, are the two newest skills that students begin from the very basic level in the IEP at DWC, as shown on their left ends which are significantly lower than the other starting points. Writing and reading proficiencies are, however, acquired with relatively good speed—their slope values are higher than the rest.

Moving to the right end of solid lines in Figure 2, most students who complete their eighth term reach or almost reach the minimum achievement rate of 80%, which is necessary for matriculation in post-secondary education, in four skills—speaking, grammar, writing, and listening, in the order of the highest average.

[5] Slope values from the equation from the highest to the lowest: writing 4.08, reading 3.52, listening 3.41, speaking 3.21, grammar 2.37, and vocabulary 1.45

Reading and vocabulary skills are coming along toward the finish line, but their linear forecast (dotted lines) indicates that it takes a few additional terms for most students to meet or to be close to the 80% achievement cutoff in reading and vocabulary. One possible approach is to increase the quantity and quality of reading and vocabulary instruction in the IEP so that these skills are mastered within a similar timeline as the rest.

Although Figure 2 illustrates different learning trends across language skills, the six skill scores do correlate. Table 2 shows statistically high positive correlations among them, having the strongest relationship between writing scores and speaking scores with their correlation coefficient 0.843 with a p-value < 0.01. This means 84.3% of variability of each skill can be predicted by another with less than one in a thousand chance of being wrong. Given that even a moderate correlation begins from the coefficient value 0.50[6], Table 2 demonstrates that the progress of each skill area affects and supports one another in the eight-week program. One's ESL scores are not to be seen separately but rather as being mutually defined, which implies that taking an oral communication course can help one to be a better reader or writer or vice versa, for instance.

[6]

Size of Correlation	Interpretation
.90 to 1.00 (−.90 to −1.00)	Very high positive (negative) correlation
.70 to .90 (−.70 to −.90)	High positive (negative) correlation
.50 to .70 (−.50 to −.70)	Moderate positive (negative) correlation
.30 to .50 (−.30 to −.50)	Low positive (negative) correlation
.00 to .30 (.00 to −.30)	negligible correlation

Pairwise correlations

Variables	(w2)	(s2)	(l2)	(g2)	(v2)	(r2)
w2	1.000					
s2	0.843***	1.000				
l2	0.674***	0.728***	1.000			
g2	0.724***	0.780***	0.580***	1.000		
v2	0.650***	0.670***	0.606***	0.650***	1.000	
r2	0.545***	0.521***	0.599***	0.463***	0.484***	1.000

Table 2. The correlation among language skills in the eight-week program

Limitations and Implications

Further research with larger and more longitudinal data sets is needed to strengthen findings in this study. Another limitation of this study is different modes of testing per skill —writing is measured by a student's hand-written essay, which is manually graded based upon the DWC IEP's own EWT rubrics. Speaking is assessed in person as well, in the format of a 10-minute interview with two normed proctors who use standard rubrics. Although both EWT and EST rubrics are informed by the Common European Framework of Reference for Languages (CEFR), there may be significantly different findings if writers type an essay and speakers record answers on the computer, which some IEPs actually do and is a common practice in commercial tests such as TOEFL iBT, IELTS, Versant, iTEP, and Duolingo. On the other hand, student scores of listening, grammar, vocabulary, and reading used in this study came from one source, the computer-based Cambridge Michigan Language Assessment. It is important to note that the reading section in CaMLA is always taken last after an almost full-hour test of other skills and students are often unable to complete the section within the given time; this means reading score data used in this study may not reflect DWC IEP students' true reading performance.

Despite these limitations, this study has led some improvements in the IEP at DWC, fulfilling its formative purpose. From Fall 2021, lower-level reading instruction hours increased from the previous two to the current three per week, which is supported by the finding that most students begin to read in English from scratch and thus would benefit from more explicit, in-class teacher intervention and facilitation. A relevant proposal to this, currently under review, is cutting oral communication hours from the current ten to eight per week, based on the finding of this study that speaking reaches the 80% benchmark, the earliest of all skills, and instead increasing reading contact hours up to 4-5 per week, while still respecting the maximum industry-standard instructional hours (23 hours/week). Another item worth discussing is selecting vocabulary textbooks for upper-level students' self-directed vocabulary practice, in hopes to increase the absolute quantity of vocabulary input in the later phase of their ESL study and expedite their vocabulary acquisition curve.

One benefit of statistical analysis that the DWC IEP notices is that it provides the groundwork not simply for brainstorming of action-plan ideas, but more importantly, for setting a priority of such ideas in the busy field of language education. It may enable a language program to break down manageable, data-driven projects into an effective timeline. Qualitative data purposefully complicate the problem with rich descriptions, while quantitative data conveniently tidy up the problem. Given the IEP director and CEA accreditation coordinator's reporting responsibility on a regular basis, numerical evidence that the program actually works and serves student needs can be a powerful addition to their report, fulfilling its summative purpose. Those student progress data also naturally contribute to a number of CEA accreditation standards that an IEP may need to respond to during its self-study (e.g. student achievement, length and structure of program of study, program development, planning, and review). Finally, it is hoped that more success and challenge stories of measuring student achievement and evaluating program effectiveness are published and shared within and for a broader intensive English program community.

References

Allen, M. J. (2004). Assessing academic programs in higher education. San Francisco: Wiley.

Brown, J. D. (2013). Teaching statistics in language testing courses. Language Assessment Quarterly, 10(3), 351-369.

California State Department of Education. (1994). Student Oral Language Observation Matrix (SOLOM). Retrieved from: https://sde.ok.gov/sites/default/files/21-22%20Final%20SOLOM%20Matrix.pdf.

Hamdoun, W. M. (2021). Academic English program evaluation of the foundation year at Saudi Petroleum Services Polytechnic College: Trainees and EFL instructors' perceptions. International Journal of Applied Linguistics & English Literature, 10(4), 24-39.

Juffs, A. (2020). Aspects of language development in an intensive English program. New York: Routledge.

Lynch, B. K. (1996). Language program evaluation: Theory and practice. Cambridge: Cambridge University Press.

Norris, J. M. (2009). Understanding and improving language education through program evaluation: Introduction to the special issue. Language Teaching Research, 13(1), 7-13.

Powers, K. (2020). Data strategy in colleges and universities: From understanding to implementation. New York: Routledge.

Scriven, M. (1967). The methodology of evaluation. In: R. Tyler, R. Gagné & M. Scriven (Eds.), Perspectives of curriculum evaluation (pp. 39-83). Chicago: Rand McNally.

Shawer, S. F. (2013). Accreditation and standards-driven program evaluation: Implications for program quality assurance and stakeholder professional development. Qual Quant, 47, 2883-2913.

Stufflebeam, D. L. (2003). The CIPP model for evaluation. International Handbook of Educational Evaluation, 31-62.