TESL Reporter

Vol. 54 (1 & 2) October 2022

ARTICLES

TIPS FOR TEACHERS

TESL Reporter

A Forum for and by Teachers of English to Speakers of Other Languages

TESL Reporter

A Forum for and by Teachers of English to Speakers of Other Languages

Vol. 54 (1 & 2)
October 2022

ISSN 0886-0661

A Publication of the English Language Teaching and Learning Program, Brigham Young University–Hawaii

TESL Reporter

Volume 54 (1 & 2) **October 2022**

CONTENTS

"Eulogy at the Passing of a Praiseworthy Professional Periodical"

Foreword to the final issue of the *TESL Reporter*

by Lynn Henrichsen

"Let's start a professional journal for TESL!" This audacious proposal, voiced casually in 1967, by William D. Conway and Alice C. Pack, faculty members at the Church College of Hawaii (CCH) may have seemed bold and ambitious. However, having just created one of the first academic degrees (a BA-TESL) in the United States in the relatively new field of teaching English as a second language, these self-assured academics were feeling bold. In 1967, there were only a few journals in the field of second language teaching—*The Modern Language Journal* (started in 1916 by the National Federation of Modern Language Teachers Associations and focused on the teaching of all modern foreign languages), *English Language Teaching* (now titled *ELT Journal*, started in 1946 by the British Council), *Language Learning* (begun at the University of Michigan in 1948 and heavily focused on research and theory), and the *English Teaching Forum* (published by the U.S. Department of State starting in 1962, but by law available only outside the United States). In 1967, two other journals oriented toward language teaching and learning were just starting up, sponsored by newly formed professional organizations: *Foreign Language Annals* (by the American Council on the Teaching of Foreign Languages [ACTFL]) and the *TESOL Quarterly* (by Teachers of English to Speakers of Other Languages [TESOL]).

Seeing a need, and with support from the English Language Institute at CCH, these two intrepid academics put together a newsletter-like publication, dubbed it the *TESL Reporter*, and in Autumn 1967 sent it out to teachers and institutions they knew were involved in teaching English as a second language in Hawaii, elsewhere in the Pacific (in the Education System of the Church of Jesus Christ of Latter-day Saints, of which the Church College of Hawaii [later BYU-Hawaii] was a part), the United States, Canada, and Asia. The "central focus" of the TESL Reporter was to be "upon the methods and problems of TESL, mostly in Hawaii and in the Pacific Basin" and it would "contain practical lesson plans, news of the ELI and BATESL programs of Church College and of other institutions, articles on language and pedagogy,…and other relevant articles of general interest" (Reporter Focuses on Hawaii, Pacific, 1967).

Printed on colored paper, the first issues of the *TESL Reporter* were only eight (6" x 9") pages in length, and initially Conway (editor) and Pack (assistant editor) wrote most of the articles themselves. They also did the typesetting and page layout, as well as the mailing address labels and bulk mailing. The first articles were mostly about TESL at CCH (a new language laboratory, the new BA-TESL degree, visits to the CCH campus by distinguished linguists), and the language training of Peace Corps volunteers at CCH. However, the first issue also had an academic-pedagogical article on teaching the [ɪ]-[i] contrast to Polynesian learners of English, as well as a discussion of cross-linguistic and cross-cultural differences that create challenges for Polynesian learners of the English language.

As the *TESL Reporter* reached the hands of TESL practitioners in its target area, academics from other institutions contributed manuscripts for publication. Ted Plaister, Gerald Dykstra, and Yao Shen (all from the University of Hawaii) contributed articles that appeared in the first few issues. For volume 1, issue 2, J. Donald Bowen of UCLA authored "Terminal Behavior and Language Teaching," which had a far-reaching effect as it was later reprinted in Kenneth Croft's *Readings on English as Second Language: For Teachers and Teacher Trainees* (1972), an anthology widely used in TESOL teacher-preparation programs of the day. Other noteworthy TESOL professionals who published articles in the TESL Reporter in its early years included Virginia French Allen, Richard Via, Nancy Arapoff, Jason B. Alter, Albert Marckwardt, Kenneth Croft, Emilio Cortez, Donna Ilyin, Judy E. Winn-Bell Olsen, Larry Smith, Peter H. Fries, Judy Gilbert, Don L. F. Nilsen, and Paul (I. S. P.) Nation. Of course, in accordance with its stated purpose, the *TESL Reporter* also included articles written by TESL faculty and students at CCH (which later became BYU-Hawaii).

As its circulation increased, the *TESL Reporter* put CCH "on the map" of the TESOL world. Its BA-TESL Program was featured in the *TESOL Quarterly* (Conway, 1969). Hearing about the CCH/BYUH TESL program, a stream of language and linguistics dignitaries visited the CCH campus over the years (including Gerald Dykstra, Robert Lado, Edward Anthony, James Alatis, Harold B. Allen, Charles C. Fries, Arthur Henry King, and Lily Wong-Fillmore). As publication of the *TESL Reporter* continued, the addition of book reviews, job announcements, conference announcements, and bibliographies increased the journal's usefulness to subscribers. It also enhanced the reputation and employability of CCH/BYUH's BA-TESL graduates, who went on to fill important positions at educational institutions throughout the Pacific and beyond.

After the *TESL Reporter's* first two years of publication, in Fall 1969, its editor, William D. Conway, left CCH, and his assistant, Alice C. Pack, became editor. Dr. Pack (despite a heavy teaching and administrative load, no workload credit for serving as editor, and minimal staff and budget) continued to nurture the fledgling publication for the next 11 years. A woman of many talents, she did everything—recruiting and reviewing manuscripts, typesetting on a Fuller Typositor, pasting up the paper galleys (using hot wax) on a light table, masking the negatives for the printer, and even sticking mailing labels on the final, printed products and bundling them by postal code with rubber bands for bulk mailing. However, in spring of 1980, Dr. Pack retired, and her assistant editor, Lynn E. Henrichsen, inherited the job. Over the next twelve years, he revised the format of the printed publication, created a new cover design, arranged for the indexing of *TESL Reporter* articles in *Language and Language Behavior Abstracts* and their inclusion in the ERIC Document database, recruited an editorial advisory board, converted from paper-based to computer-based typesetting and page layout, acquired an ISSN (International Standard Serial Number) for the journal, and increased the *TESL Reporter's* circulation and sphere of influence. By its fifteenth anniversary, the journal had a circulation of nearly 3,000 subscribers in over 60 countries. By 2010, that number had grown to over 100 countries.

When Henrichsen left BYU-Hawaii in 1992, Mark O. James became editor. To appeal to a wider range of potential contributors, he further defined manuscript categories: Full-length manuscripts, Teaching Tips, and Reviews (of books, sites, apps and software). He also changed the frequency of publication to twice a year and increased the size of each issue to 80-100 pages. Around the same time, to offset increasing printing and mailing costs, annual subscription rates for the

TESL Reporter were increased to US$6 for U.S. residents and institutions, but in harmony with its initial mission the periodical was still distributed on a complimentary basis to subscribers outside the United States. In ensuing years, other editors (Maureen S. Andrade, Mark Wolfersberger, Nancy Tarawhiti, and most recently Austin Pack and Jefferey Maloney) have made similar improvements, such as switching from paper-based to online publication, which made instantaneous worldwide distribution of the *TESL Reporter* possible and also saved BYUH thousands of dollars in international postage fees.

Because it was initially sent to subscribers by request and published free of charge—in contrast to the steep subscription fees of most academic journals—the *TESL Reporter* was able to reach a large and deserving audience. It went not just to academics at high powered universities, but to teachers in developing countries and libraries at struggling schools. For many of these recipients, the *TESL Reporter* was the only publication that connected them with the broader, professional world. Its short, practical, pedagogically oriented articles were especially appreciated. Occasionally subscribers would write and express their thanks. For example, a grateful subscriber once wrote, "The *TESL Reporter* is the only professional journal that I read from cover to cover when I receive it. Because of its small size, it fits in my pocket, so I can read it on the bus to work, and the short, practical articles keep my interest."

On a personal, autobiographical note, I owe the start of my own academic publishing career to the *TESL Reporter*. In 1975, as a recent MA graduate traveling to a new job with the American Samoa Department of Education, I stopped in Hawaii enroute and visited Laie. There I was generously given a stack of back issues of the *TESL Reporter* to take with me to Pago Pago. As I settled in, I put the colorful little journals on my desk, and every morning, I would open one up and read an article or two. The better ones informed and inspired me in the performance of my professional duties and kept me connected with the broader world of English Language Teaching. The articles of lesser quality were also inspirational in a different way. Some of them made me think, "I could do better than that." Following up on that thought, I wrote up some lessons I had learned from my experience in my previous job managing a language laboratory. I summoned my courage, mailed my manuscript off to the *TESL Reporter*, and was thrilled when in short order I received an acceptance letter from the editor and then several, bright green physical copies of the *TESL Reporter* with my article on the front page (Henrichsen, 1976)! That experience helped me overcome my fear of failure and launched my academic publishing career. In retrospect, 46 years later, few of the 13 books, 36 book chapters, 3 monographs, and 78 other articles I have authored have been able to match the thrill I received from publishing my first article with the *TESL Reporter*. I like to think that the *TESL Reporter* played similar, beneficial roles in the careers of many other TESL/TEFL professionals.

Now, 55 years after it began, the *TESL Reporter* will cease publication. In light of the many dozens of highly professional (and specialized) academic journals that exist in the field of English language teaching, as well as the abundance of websites devoted to TESOL practices and resources, the need that led to the *TESL Reporter's* beginning no longer exists. Institutional priorities and faculty workloads have also contributed to the decision to stop producing the *TESL Reporter*. While a part of me feels sad about that, a much greater part of me is proud to have been part of this publication and its influence on the TESOL profession and the world. It definitely accomplished its noble professional purpose and did so unselfishly and well. That is an enviable obituary. Most of us will be happy if the same can be said about us and our careers.

References

Alice Pack departs. (1980). *TESL Reporter, 13* (4), 72-73, 80. http://ojs-dev.byuh.edu/index.php/Issue1/article/view/891/847

Conway, W. D. (1969). The undergraduate major in TESL, *TESOL Quarterly, 3* (1), 55-59. https://doi.org/10.2307/3586043

Croft, K. J. (1972). *Readings on English as a second language*. Boston: Little, Brown.

Henrichsen, L. E. (1976). Is the language laboratory dead? *TESL Reporter, 9* (4), 1-2, 16-17. http://ojs-dev.byuh.edu/index.php/Issue1/article/view/718/692

Reporter Focuses on Hawaii, Pacific. (1967). *TESL Reporter 1*(1), 1-2. http://ojs-dev.byuh.edu/index.php/Issue1/article/view/374/355

Lessons Learned from a Career in TESOL

Neil J Anderson, Brigham Young University–Hawaii, Laie, Hawaii, USA

Editor's note

This article is an adaptation of the 11th annual Alice Pack lecture delivered by Dr. Neil J Anderson at Brigham Young University Hawaii on June 3rd , 2021. In this article Dr. Anderson reflects on 41 years of experience of teaching English language learners and English teachers and shares three major lessons with TESOL professionals. Professor Anderson is a recipient of the prestigious TESOL International Association James Alatis Service Award (2014), and was recognized by TESOL International Association in 2016 as one of the 50 individuals who has made a significant contribution to the profession of teaching English to speakers of other languages.

Introduction

I am now within weeks of retiring from my position as a university professor. The past 41 years have provided rich and rewarding opportunities to engage in teaching English to speakers of other languages and to train teachers entering the profession. I have had the opportunity to teach at three universities in the United States and I have had extended periods of time outside of the US working and teaching in refugee language programs in Southeast Asia, and with language learners and teachers in Costa Rica and Guatemala. I have also had the opportunity of interacting with teachers in over 50 countries as part of speaking invitations to teacher conferences, training programs through the U.S. Department of State English Language Programs, and publisher workshop for books I have published.

Before I retire, I want to step back from 41 years as a TESOL professional and reflect on lessons learned. There is one overarching concept that ties together the major lessons that I have learned over my career. That concept is relationships. In his 2019 book *The second mountain: The quest for a moral life,* David Brooks emphasizes the vital role that relationships play in our lives. "As adults, we measure our lives by the quality of our relationships and the quality of our service to those relationships. Life is a qualitative endeavor, not a quantitative one. It's not how many, but how thick and how deep. Defining what a quality relationship looks like is a central task of any moral ecology" (pp. 300-301).

The relationships I have built with colleagues and students have made all the difference in what I have learned. Without these relationships, I know that I would not have lasted 41 years as a TESOL professional.

I can summarize hundreds of relationship experiences into three major lessons: first, expect the unexpected; second, foster a culture of collaboration; and third, connect your scholarship to your teaching and your teaching to your scholarship. Each of these three relationship lessons has resulted in forming and building lasting friendships with others that have molded me into a very different person today than I was 41 years ago. I am confident that the lessons I have learned can benefit other TESOL professionals, regardless of the stage at which you are currently at within the profession.

Expect the Unexpected

For some reason, we often believe that we are in complete control of the events that occur in our lives. We believe that the private decisions we make day-by-day shape us into the person we see ourselves becoming. I have come to understand that we must expect the unexpected. Many of the good things that have occurred in my professional life have come unexpectedly. These unexpected experiences have often come as challenges that have caused me to rely on key relationships to overcome them.

One way that I have come to look at the idea of expecting the unexpected is through the phrase *question all assumptions.* Quinn (2012) introduced the concept of "challenging your normal assumptions" (p. 5). His work led me to the creation of a phrase that has guided me in most activities and decisions in the past seven years: *question all assumptions.* Quinn asserts that deep change is a fundamentally different process because it requires people to develop new expectations. As people experience deep change, they move from their old assumptions to a new set of assumptions. They start to see, feel, and think differently. (p. 9)

I have not always questioned my assumptions. This is a lesson that emerged later in my career but is one that I wish I had understood early in my career. In retrospect, this lesson has provided me with opportunities to grow and develop as a TESOL professional.

Allow me to provide an example from my career. I completed my MA degree in December 1980. My first full-time teaching position was at the newly organized English Language Center (ELC) at BYU Provo. The program designated a limited number of three-year non-renewable teaching positions for newly graduated students from the MA program. The rationale for these positions was to provide full-time employment that could serve as a stepping stone to further employment and education.

Within the first few weeks of my full-time employment, I knew that I wanted to earn a PhD degree in order to train future TESOL professionals. I set my sights on the PhD program at the University of California at Los Angeles (UCLA). I was so focused on earning my PhD from UCLA that it never occurred to me that I might be rejected. In April 1984 I received the long-awaited letter from the admissions office at UCLA. I was devastated when I read that I had not been accepted. My world suddenly fell apart. I had no backup plan. It took a few weeks for me to sort through what I should do. Through very careful guidance from mentors, individuals with whom I had built relationships, I applied for the PhD program at the University of Texas at Austin (UT Austin) and I applied for a position with the Center for Applied Linguistics (CAL) at the Refugee Service Center based in Manila, Philippines. I was accepted to the PhD program at UT Austin and I was offered a position with CAL as the Program Associate for Measurement and Evaluation. Although I did not fully understand at the time, I was learning to expect the unexpected.

From this experience I also learned that in life we want options. I should never have assumed that I would be accepted at UCLA. I should never have applied to a single program. Because I applied to grad school at UT Austin and the position at CAL and received positive results from each, I was then in a position to make an informed decision. I accepted the position with CAL and I petitioned UT Austin to delay my acceptance for two years. That allowed me to have extremely positive experiences working at refugee programs in the Philippines, Thailand, and Indonesia. I was in a much stronger position to enter the PhD program at UT Austin and successfully complete the course work and dissertation within three years.

My rejection from UCLA provided two professional options that I never expected. Working for CAL provided an opportunity for our family to live in the Philippines. I was able to travel multiple times to refugee programs in the Philippines, Thailand, and Indonesia followed by studying in a PhD program that ultimately was a better fit for me and my family. I built collaborative associations with colleagues at CAL and at UT Austin that served me well years later. These were completely unexcepted events for me and my family. Expect the unexpected!

Foster a Culture of Collaboration

As a language teaching professional, I have had many opportunities to collaborate with others. Most of the collaborative opportunities came to me unexpectedly. Mercer and Dörnyei (2020) point out the value of building a culture of collaboration in the language classroom. I see it equally beneficial to build a culture of collaboration for your academic career.

One principle that I learned as a student in the MA program at BYU was to be actively engaged in professional associations. I joined TESOL International Association while still a graduate student and I joined Intermountain TESOL (I-TESOL), the affiliate of TESOL for professionals living in Utah and Idaho. Early in my membership in I-TESOL I met MaryAnn Christison (MAC), then a faculty member at Snow College in Ephraim, Utah. When MAC was President of I-TESOL she invited me to serve as the treasurer of the association. This invitation resulted in a friendship and collaboration that has lasted for years.

Early in my service opportunity with I-TESOL I met Darlene Larson who was serving as the 17th President of TESOL International Association (1982-1983). As a result of meeting her, she appointed me to serve on the first TESOL Awards Committee. This professional service opportunity allowed me to meet other professionals and to review the award applications. I started to learn what constitutes a strong application for an award and how to evaluate award applications.

JoAnn (Jodi) Crandall served as the Chair of this committee. Jodi was also the Vice President of CAL. She traveled to the Philippines shortly after I started my employment and I was able to strengthen my collaboration with her as I learned how to serve within TESOL and as an employee of CAL.

In 1993 MAC had been elected to the TESOL International Association Board of Directors as the convention program chair for the 1995 convention that was to be held in Long Beach, California, USA. She invited me to serve as her Associate Program Chair. My collaboration with MAC provided me with an opportunity to meet many more TESOL professionals from around the world that I otherwise would not have met. Serving as MAC's Associate Convention Chair also introduced me to the workings of the TESOL Board of Directors and I knew that I wanted to put my name forward for leadership opportunities within TESOL. I was elected by the membership of TESOL to serve as the Convention Program Chair for the 1998 convention that was held in Seattle, Washington, USA. After serving in that role from 1997-1999, I was elected by the membership to serve as TESOL President. That opportunity allowed me to serve as President-Elect (2000-2001), President (2001-2002) and then as Past President (2002-2004).

During the years of service on the TESOL Boards of Directors I had the opportunity of building collaborative associations with women and men who served in the role as TESOL President. Donald Freeman, Joy Reid, Denise Murray, MaryAnn Christison, Kathi Bailey, David Nunan, and Barb Schwarte have become dear friends, not just professional colleagues.

Another collaboration that was fruitful was getting to know G. Richard Tucker. Dick was the President of CAL during my employment in the refugee program. Later Dick and I served on Board of Directors for The International Research Foundation (TIRF). Dick played a significant role, along with Donald Freeman, in convincing me to serve as the Chair of TIRF (2004-2006).

Through my employment at CAL, I met Rebecca Oxford. Rebecca was employed at the Washington D.C. CAL office. She was one of my primary contacts in DC for analyzing the language placement and progress data we gathered at the refugee camps. Rebecca and I co-authored a scholarly publication on the impact of learning styles of second language learners (Oxford & Anderson, 1995). We were able to host Rebecca at BYU Hawaii a few years ago when she came to speak at Hawaii TESOL.

In February 2001 I was invited to participate in a US Department of State English Language Program opportunity as a Specialist to work with the Faculty of Education at Suez Canal University in Suez, Egypt. The focus on the training was on designing self-assessment tasks for EFL students at Suez Canal University. The training was in two parts. I met with the teachers in February and then I returned to the US allowing them to design and develop some self-assessment tasks. I was to return in September. My flight was schedule for September 13, 2001. Now pause and think. What was happening in the world in September 2001? Because of the terrorist attack on the Twin Towers in New York and other locations on September 11, all flights in the US were grounded. I do not remember the exact date, but as soon as flights reopened, I was on a plane to Egypt to meet with my colleagues at Suez Canal University. Many people were concerned about me traveling to the Middle East at such a point in time. I arrived in Suez and entered the training room. As I entered with my hosts, my Egyptian colleagues stood and applauded. I was overcome with emotion. I asked why they were applauding. One teacher stepped forward and said that they were sure that I would not return to complete the training with them. They were sure that the media reports and the terrorist attacks in the US would discourage me from returning to meet with them. I shared with them that many people were concerned about my return, but that I was not. To those that were concerned I said, “I know my Muslim sisters and brothers will protect me from any danger while I am in Egypt.” That brought another round of applause and emotional embraces that I will always remember.

I have also been fortunate enough to have two Fulbright teaching and researching fellowships: Costa Rica (2002-2003) and Guatemala (2008-2009). I was interested in going to Costa Rica for multiple reasons. One, both my wife Kathy and I speak Spanish and I wanted an opportunity to live in Central America and be surrounded by Spanish. Second, while I was on the faculty at Ohio University in Athens, Ohio I taught several students from Costa Rica. Because of those students, in 1993 I was invited to as an English Language Specialist to conduct two weeks of training for teachers. That first visit to Costa Rica resulted in multiple returned visits. Two of my former students at Ohio University, Evelyn and Hellen Aguilar Murrillo are twins. They hosted me on most of those visits. Another former student Maria Eugenia Flores, was the Academic Director at the Centro Cultural Costarrisense Norteamericano (CCCN) and her Associate Academic Director was another former student Alexandra Esquivel. I knew that spending a year in Costa Rica would provide powerful collaborations.

It was at a faculty research meeting at the University of Costa Rica that the concepts of developing depth in the development of teaching the language skills was conceived. I have continued to refine my thinking about how to appropriately integrate the language skills of listening, speaking, reading, and writing and at the same time develop depth in each of the skills. I have a recent publication in the TESOL Encyclopedia of English Language Teaching (Anderson, 2019) on how to develop depth in the development of reading while integrating the other language skills.

I served as a missionary in Guatemala and El Salvador in the mid-1970s and returning to Guatemala as a Fulbrighter was particularly rewarding for me. I was able to work with teachers at the Centro de Aprendizaje de Lenguas Universidad de San Carlos (CALUSAC). The term motivational moments was conceived while working with and training teachers in Guatemala. We met weekly to discuss how to make motivation a more explicit component within language classrooms. The teachers became overwhelmed with all of the theories and activities related to motivation. One day one of the teachers stop our discussion and said, "Neil, we cannot possibly do all of this. We have to teach the language." I started to panic because I could see that the focus and purpose of my being in Guatemala as a Fulbrighter was slipping away. This is one time that I know I was inspired as to how to respond. I paused, looked the teachers in the eyes, and said that we needed to shift our thinking about the theories and activities related to building motivation as opportunities to weave in motivational moments into our lesson planning. On the spot I told the teachers that a motivational moment should last no more than one minute. With this explanation the entire attitude of the teachers changed. They unanimously agreed that they could easily weave four motivational moments into each lesson plan. We then started practicing the delivery of those motivational moments in order to "create the basic motivational conditions, generate initial motivation, maintain and protect motivation, and encourage positive retrospective self-evaluation (Dörnyei, 2001, p. 29). My experience in Guatemala ended with strong collaborative associations and teachers that learned how to weave motivational moments into their teaching.

While on the faculty at BYU Provo I had the opportunity to interact with extremely motivated and talented students. It was in that context that I first met Rob McCollum, Rebecca Roberts, Nancy Tarawhiti, and Mark Wolfersberger. Today I actually do not think of Mark, Becky, and Nancy as "former graduate students" but rather as colleagues.

Mark initiated the discussion with me about teaching here at BYU–Hawaii. In 2014 I shared with Mark that I was questioning the assumption that I would remain at BYU until retirement. I wanted to do something different for the final years of my academic career. I had applied for two positions. He asked if I would consider applying for a position at BYU–Hawaii. That invitation provided a rich and interesting discussion with him and with the then Academic Vice President, Max Checketts. After lengthy discussions with both Mark and Max, I submitted my application. Kathy and I were flying home from one job interview when we received the call that we were invited to come to Hawaii for an interview. I resigned from my position on the faculty at BYU Provo and we moved to Laie in 2014.

Fostering a culture of collaboration has provided many rich and rewarding relationships.

Connect your Scholarship to your Teaching and Your Teaching to your Scholarship

Early in my career I became interested in the skill area of reading. I decided that this was the skill area that I wanted to explore and develop expertise in so that I could improve my teaching of reading fluency and reading strategies. My desire came as a result of seeing students in my ESL classrooms struggle with slow reading and lacking strategic reading skills. I decided early in my career that I would connect scholarship to challenges I saw in my teaching. My dissertation examined the reading strategies that L2 readers use while engaged in reading comprehension tests versus reading academic textbook material (Anderson, 1989).

In 1984 TESOL International Association held its annual convention in Houston Texas. I attended the annual Reading Research Colloquium. That was my first formal introduction to David Eskey, Patricia Carrell, Andrew Cohen, and Bill Grabe. As the colloquium drew to a close, I took the initiative to approach the organizers and share with them my developing interest in reading fluency. They invited me to join them for dinner that evening. That dinner initiated friendships with these colleagues that continue today. Andrew ended up serving as an external committee member for my dissertation work at UT Austin.

I presented my dissertation research at the 11th annual Language Testing Research Colloquium in San Antonio, TX, USA in 1989. Andrew was in the audience. After the presentation, Lyle Backman and Kyle Perkins approached us and asked detailed questions about my data. The editor of *Language Testing* also approached us during that conversation. The conversation and collaboration resulted in a joint presentation the next year at the 12th annual Language Testing Research Colloquium in San Francisco, CA, USA and in a 1991 publication in *Language Testing* (Anderson, Bachman, Perkins, & Cohen, 1991). By connecting my scholarship to my teaching and my teaching to my scholarship I also had the opportunity of making meaningful relationships with other professionals.

In 2000, I was invited to deliver a plenary address at the Millennium Malaysia International Conference on English Language Teaching in Malacca, Malaysia. I spoke on developing metacognitive skills in second language learners. I have found that a conference presentation is the perfect starting point for developing a publication to submit for a peer reviewed publication. While at this conference I was able to reconnect with a former student I had taught while at Ohio University, Josephine Ratnam-Wee. The relationships established with students during their time in my classes have led to multiple opportunities to keep those relationships alive.

While at the conference I was able to interact with David Nunan, who at that time was serving as the president of TESOL International Association. I had accepted an invitation from him to contribute a chapter to a book that he was in the early stages of editing Practical English Language Teaching. I had shared with David an initial abstract for my chapter which presented a framework for teaching reading. The ACTIVE skills for reading framework is one that I had developed as a way to approach my teaching of reading in the classroom. At the conference David introduced me to Ian Martin, an editor with Heinle & Heinle Publisher, now National Geographic Learning. While Ian, David, and I were relaxing that evening after the conference David asked me to share with Ian my framework. David then told Ian that he should get a contract with me to publish a textbook series by this title. Within six months I had a contract to write four textbooks: ACTIVE Skills for Reading, Books 1-4 (Anderson, 2002/2003). When the series went into its second edition for publication in 2007/2008 (Anderson, 2007/2008), we added an Introduction level book. In 2013/2014, the series then went on to a third edition (Anderson 2013/2014). When I first developed the framework at the ELC in 1981, I had no idea that my thinking would result in multiple conference presentations and three editions of a textbook series. Notice with this example that it was my relationship with David Nunan that provided the introduction to Ian Martin which led to my scholarship resulting in a textbook series.

My scholarship has always emerged from issues that I have faced in language classrooms. Explicitly connecting my scholarship with my teaching and my teaching to my scholarship has resulted in eight books, 29 book chapters, 27 journal articles, and three textbook series.

Conclusions

Relationships are core to our lives. Those relationships need to be rich, purposeful, and meaningful. We do not know how we ultimately might be influenced by others or how we might influence them. The relationships I have developed over the past 41 years have prepared me to expect the unexpected, foster a culture of collaboration, and connect my scholarship to my teaching and my teaching to my scholarship.

I am confident that other TESOL professionals can benefit from these same principles. Then when you reach the point that you have spent 40 years in the profession, I hope you can reflect back on the relationships you have established and retire knowing that you have accomplished good.

References

Anderson, N. J. (1989). *Reading comprehension tests versus academic reading: What are second language readers doing?* [Unpublished doctoral dissertation]. The University of Texas at Austin.

Anderson, N. J., Bachman, L., Perkins, K., & Cohen, A. D. (1991). An exploratory study into the construct validity of a reading comprehension test: Triangulation of data sources. *Language Testing, 8*, 41-66.

Anderson, N. J. (2000). Developing metacognitive skills in second language learners. Plenary address at the Millennium Malaysia International Conference on English Language Teaching, Malacca, Malaysia.

Anderson, N. J. (2002/2003). *ACTIVE Skills for Reading*, Books 1-4. Heinle Publishers.

Anderson, N. J. (2007/2008). *ACTIVE Skills for Reading*, (2nd ed.), Books Intro, 1-4. Heinle Publishers.

Anderson, N. J. (2013/2014). *ACTIVE Skills for Reading*, (3rd ed.), Books Intro, 1-4. Heinle Publishers.

Anderson, N. J. (2019). Integration with other language skills. In *The TESOL Encyclopedia of English Language Teaching*. Edited by John I. Liontas (Project Editor: Margo DelliCarpini; Volume Editor: Gloria Park). John Wiley & Sons, Inc.

Brooks, D. (2019). *The second mountain: The quest for a moral life*. Random House.

Dörnyei, Z. (2001). *Teaching and researching motivation*. Harlow: Longman.

Mercer, S., & Dörnyei, D. (2020). *Engaging language learners in contemporary classrooms*. Cambridge.

Oxford, R. L. & Anderson, N. J. (1995). State of the Art: Crosscultural View of Learning Styles. *Language Teaching, 28*, 201-215.

Quinn, R. E. (2012). *The deep change field guide: A personal course to discovering the leader within*. John Wiley & Sons.

About the Author

Neil J Anderson is a Professor Emeritus of English Language Teaching and Learning at Brigham Young University–Hawaii. In 2016, he was recognized by TESOL International Association as one of the 50 individuals who has made a significant contribution to the profession of teaching English to speakers of other languages.

A Vocabulary Deep Dive into the National Geographic Textbook *Reading Explorer 4*

Suzanne H. Rice, Brigham Young University, Provo, Utah, USA
Grant Eckstein, Brigham Young University, Provo, Utah, USA

Abstract

This brief report investigates the vocabulary used in the 24 articles of National Geographic's *Reading Explorer 4* textbook and compares them according to the CEFR and ACTFL frameworks whilst examining words featured in each chapter's highlighted vocabulary list. The purpose is to juxtapose the stated purpose of a textbook and its utility in an intensive English program's intermediate high reading class. A lexical analysis showed that while the chapter word lists feature advanced vocabulary, they only average 20% of AWL words. Articles cover 57% of the AWL in the book but with little repetition, and the articles in each chapter do not increase in difficulty. A pedagogical discussion follows as to how teachers can supplement and support the vocabulary needs of their students.

Keywords: *vocabulary, AWL, ACTFL, CEFR, Reading Explorer*

Introduction

Reading teachers, particularly those new in the English as a second language (ESL) profession, may adopt a particular reading textbook assuming it is a good match for student learning because publishers advertise it as such. If not critically examined, teachers may assume that a textbook would recycle important academic words, introduce increasingly difficult vocabulary in each successive chapter, and would align with established performance proficiency benchmarks from organizations such as the Common European Framework of Reference for Languages (CEFR) or the American Council on the Teaching of Foreign Languages (ACTFL). While these assumptions might be correct, the alignment between textbooks and students needs is often not so straightforward. This brief report illustrates the lexical alignment analysis one teacher performed who adopted National Geographic's *Reading Explorer 4* textbook for a mid-intermediate ESL course.

Background

A major goal of reading teachers in English language instruction courses is to assist students in improving their reading ability and vocabulary knowledge. University professors report that reading skills are more necessary than writing or speaking skills in college (Hartshorn et al., 2017, 2019), and vocabulary researchers emphasize the need for students to develop a working knowledge of 95% of words in order to comprehend the text (Gardner, 2013). While extensive reading is one way to foster reading development and vocabulary exposure, Grabe (2009) argues that language development courses should provide practice in both learning to read and reading to learn. Preparatory materials which are not appropriately selected can result in learner frustration that can lead to student failure (Lynn, 2021). Thus it is critical to understand the appropriateness of texts prior to adopting them. One measure of appropriateness is the alignment of the vocabulary with an established list. In this sense, it is the connection between students' vocabulary needs and the vocabulary presented in course reading materials.

Vocabulary Needs

The *English Profile* program within the CEFR identifies lexical items that learners should know at each of the six CEFR levels (Cambridge, n.d.). The six levels are A1, A2, B1, B2, C1, and C2, and they represent levels of proficiency which will be discussed in more detail in the Methods section below. The CEFR B1 level manual explains that the CEFR is not in favor of teaching specific vocabulary but instead prepares learners for benchmark workplace or social capabilities (Van Ek & Trim,1998). Schools, publishers, and assessors in the United States tend to draw from the Academic Word List (AWL) (https://www.wgtn.ac.nz/lals/resources/academicwordlist) which contains 570 word families that are not on the General Service List and make up 10% of the vocabulary used in academic texts (Coxhead, 2000). Because ACTFL does not suggest or propose any leveling of vocabulary, the AWL has become the de facto vocabulary teaching plan for students preparing for academic exams.

In order to demonstrate vocabulary alignment, a reading textbook should introduce vocabulary consistent with CEFR or AWL vocabulary levels. That is, lexical items that are more common (i.e., more frequently used) in English should be introduced before items that are less common. For example, A1 vocabulary items should be introduced before A2, B1, B2, and so forth. Similarly, lexical items on the first of the 10 sublists of the AWL should be introduced before items on list 2 and so forth. Ideally texts within a textbook should move from more to less common vocabulary as the book progresses. Vocabulary alignment can be measured by comparing the lexis in reading texts to the established CEFR and AWL lists as well as general frequency lists of English words to determine whether vocabulary becomes increasingly difficult (i.e., less common) as the textbook progresses. This study examined vocabulary alignment of *Reading Explorer 4* by asking the following research questions:

1. To what extent do the articles in *Reading Explorer 4* introduce increasingly more difficult vocabulary?
2. To what extent do the articles and highlighted vocabulary lists in *Reading Explorer 4* cover and recycle words on the AWL?
3. To what extent do the AWL words correspond to the CEFR level ratings of vocabulary words in the articles?

Methods

Materials

The material used in this analysis were the *Reading Explorer 4* textbook, which is the fifth book in a six-book series designed within the Common European Framework of Reference for Languages (CEFR). It is comprised of 24 real-world articles as well as 20 lists of key vocabulary to assist with readings. The book has a level rating of B2/C1 - intermediate high. At the B2 level a learner "can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialisation" and at C1 a learner "can understand a wide range of demanding, longer texts, and recognise implicit meaning" (Cambridge, 2013, p. 5). This is comparable to the ACTFL standards for intermediate high which state that a learner "can usually follow the main message in various time frames in straightforward, and sometimes descriptive, paragraph length informational [and fictional] texts" (NCSSFL-ACTFL, 2017b, p. 2) and at advanced low a learner "can identify the underlying message and some supporting details across major time frames in descriptive informational [and fictional] texts" (NCSSFL-ACTFL, 2017a, p. 2). Both frameworks seem to agree on learners being able to comprehend the main idea, details, and some implicit meanings found in fictional and nonfictional texts of varying lengths.

Procedure

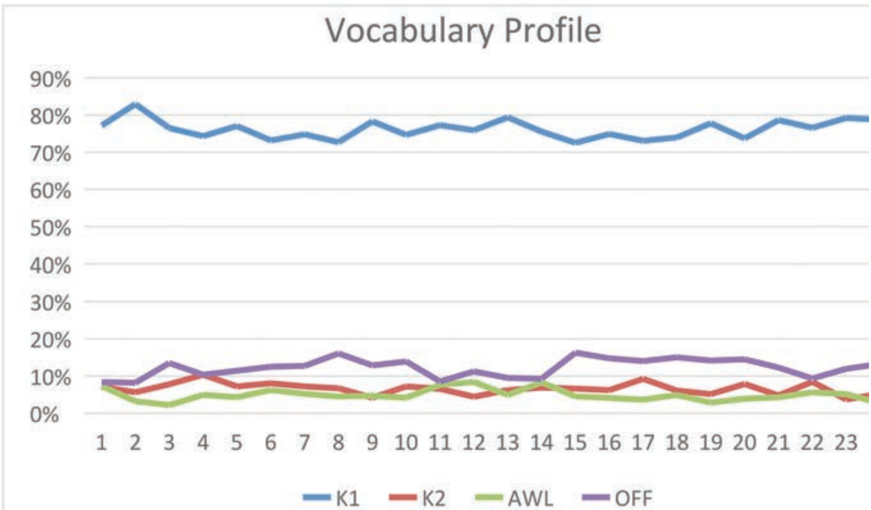
Each of the 24 articles were converted to text files and run through the Web VP Classic program on LexTutor.ca. The percentages were calculated for the first and second 1000 words of English, the AWL, and off-list words in the article. Then each file was run through the Profiler on VocabKitchen.com to calculate the CEFR vocabulary level ratings and percentages for each word level. The same process was performed for the 20 highlighted vocabulary word lists for each chapter in the book. All the articles were run in the Range and Compleat Web VP programs to get a complete look at which AWL words were included and the frequency of occurrence. A list was compiled with each of the 570 AWL words, their sublist, CEFR level rating, and frequency in the textbook.

Results

Vocabulary Progression

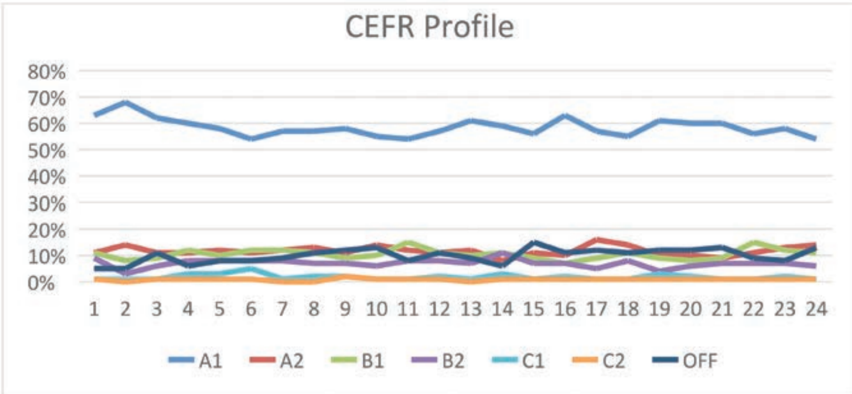
The first research question asked whether *Reading Explorer 4* introduced increasingly more difficult vocabulary as the book progressed. Results in Figures 1 and 2 show that the articles all used approximately the same percentage of each level of words from the various profiles without any substantial change from article to article. In fact, as seen in Figure 1 the very first article, labeled as 1, contained 7% AWL words while the last article, labeled as 24, only contained 3%. There are 12 chapters with 2 readings per chapter in the book. We expected to see the difficulty increase with each reading, but as can be seen in Figures 1 and 2, that is not the case. Conversely, as seen in Figure 2, the first article contained 63% A1 words on the CEFR profile compared to 58% on the final article and an increase of 8% from off-list words, indicating an increase in vocabulary difficulty; however, there was no change in C1 or C2 word percentages indicating that the change was mostly due to uncommon or off-list words appearing in the final article.

Figure 1. Percent of words in each article appearing on various lists according to Vocabulary profile



Note. K1 = the 1000 most frequent words in English according to the BNC COCA; K2 = the second most common thousand words in English; AWL = Academic Word List; OFF = all words not appearing on the K1, K2, or AWL lists. 1 through 24 = readings in the National Geographic text corresponding to two passages per chapter (e.g., 1 and 2 = first and second reading in chapter 1).

Figure 2. Percent of words in each article according to CEFR profile



Note. A1 - C2 = CEFR proficiency bands (Cambridge University Press, n.d.)

Coverage of AWL

The second research question asked to what extent the articles and highlighted vocabulary lists in *Reading Explorer 4* covered and recycled words on the AWL. Results showed that the words from the articles in the *Reading Explorer 4* textbook covered 57% of the AWL. However, frequency data from the Range program showed that AWL words had limited recycling in successive chapters: only 25% of AWL words were used in more than one article (Appendix A shows a breakdown of repeated AWL words). Academic words were not introduced based on frequency as every article contained words from sublists 1-9. Moreover, only 60 of the 240 items on the highlighted vocabulary lists came from the AWL, covering only 10% of the list. On average 4 of the 20 words from each chapter came from the AWL and none were repeated. In contrast, nearly all the highlighted words came from levels B2-C2 of the CEFR.

AWL Overlap with CEFR

The final research question asked whether the AWL words correspond to CEFR level vocabulary words in the articles. Results showed that the CEFR and AWL sublists overlapped inasmuch as nearly all AWL words (90%) were included in a CEFR level from A1-C2. Further, the CEFR levels sampled from the AWL sublists in a roughly linear pattern where lower CEFR lists also sampled from lower AWL sublists as illustrated in Table 1. The majority of AWL words used in *Reading Explorer 4* (52%) were also B2 words, indicating that vocabulary used in the text largely corresponded to the language level of target learners.

Table 1. AWL word type by CEFR level

	AWL Sublists									
CEFR List	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
A1	0	2	0	1	0	0	0	0	0	0
A2	2	3	2	2	2	1	2	2	1	2
B1	17	12	12	17	5	12	17	8	6	2
B2	33	31	26	19	29	18	33	16	13	6
C1	7	7	9	11	14	16	7	15	18	6
C2	0	1	5	6	5	6	0	9	13	10
Off	1	3	6	4	5	6	5	1	8	4

Conclusion

This analysis of *Reading Explorer 4* showed that articles did not use increasingly difficult words as the book progressed. This is especially true for AWL words which seemed to be chosen at random, highlighted capriciously throughout the text, and introduced only once. CEFR vocabulary levels demonstrated a very slight increase in difficulty as the book progressed, but that increase in difficulty was related to the inclusion of off-list vocabulary rather than higher levels of C1 or C2 lexical items.

The lack of alignment of vocabulary with AWL and CEFR lists in *Reading Explorer 4* has pros and cons. The cons include the fact that the book does not offer a progressive lexical challenge to students, so those who study from it will not necessarily be aided in developing increasing academic vocabulary. Moreover, it is unusual for teachers to cover every chapter of a textbook in a semester, so exposure to academic words may be limited by this factor. Also, the academic words are not repeated from article to article leaving little chance for repetition or spaced retrieval. Students would need supplementation by the teacher to get repeated practice with academic words. On the other hand, a benefit is that students can read articles out of order and according to their interest without affecting the progression of the book. Additionally, the featured words in each chapter are advanced vocabulary and bring a real world context that is interesting for students and may lead to increased sensitization and comprehension of these items. Readers will nonetheless be exposed to a variety of B2 vocabulary words, which aligns with the CEFR orientation of the text.

The vocabulary analysis revealed that *Reading Explorer 4* was not focused on the repetition of AWL words, but it was concerned with introducing level appropriate words that would help learners to understand the main idea and determine implicit meaning from text. Knowing this, teachers can add their own academic vocabulary word activities to encourage vocabulary progress for different purposes such as preparing students for academic writing or testing in non-European contexts.

The analysis in this study further demonstrates one simple and inexpensive way for teachers to analyze and understand the vocabulary coverage of textbooks they adopt. Teachers can enter electronic text into analyzers on websites such as LexTutor.ca and VocabKitchen.com to quickly evaluate vocabulary alignment across chapters in a book or between books they are considering adopting. In addition to the websites used in this analysis, other analysis tools such as wordandphrase.info, which creates frequency-based word lists and uses the Corpus of Contemporary American English as its underlying source, can allow both students and teachers to examine vocabulary coverage and conduct further in-depth collocation and meaning searches of individual words and phrases.

Although the specific findings related to vocabulary alignment of *Reading Explorer 4* are not generalizable to other ESL textbooks, the principles of vocabulary analysis are. Further investigations should be conducted on textbooks in this and other series. Hopefully the process and findings in this study will inspire teachers in their own textbook analyses.

References

Cambridge University Press. (n.d.) *English profile: Introducing the CEFR for English* [Booklet]. http://englishprofile.org/images/pdf/eng_pro_information_booklet.pdf

Cambridge University Press. (2013). *Introductory guide to the common European framework of reference (CEFR) for English language teachers* [Booklet]. http://englishprofile.org/images/pdf/GuideToCEFR.pdf

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213-238. https://www-jstororg.erl.lib.byu.edu/stable/3587951?seq=4#metadata_info_tab_contents

Gardner, D. (2013). *Exploring vocabulary: Language in action*. Routledge.

Grabe, W. (2009). *Reading in a second language: Moving from theory to practice.* Cambridge.

Hartshorn, K.J., Evans, N.W., Egbert, J., & Johnson, A. (2017). Disciplinespecific reading expectation and challenges for ESL learners in US universities. *Reading in a Foreign Language, 29*(1), 36–60.

Hartshorn, K., Hart, J. M., & McMurry, B. L. (2019). Comparing language skill priorities among TESOL faculty and ESL students bound for Englishmedium universities. *TESOL Journal*, 10(3), e00438.

Lynn, E. M. (2021). Comparing lexical bundle use in EAP reading textbooks to lower-division university textbooks. *Pedagogical Linguistics*. Advanced online publication. https://doi.org/10.1075/pl.21001.lyn

MacIntyre, P. & Boholke, D. (2014). *Reading Explorer 4* (3rd ed.). National Geographic/(ELT).

NCSSFL-ACTFL. (2017a). *Can-do statements: Performance indicators for language learners: Advanced* [Booklet]. https://www.actfl.org/sites/default/files/candos/Advanced%20Can-Do_Statements.pdf

NCSSFL-ACTFL. (2017b). *Can-do statements: Performance indicators for language learners:* Intermediate [Booklet]. https://www.actfl.org/sites/default/files/candos/Intermediate%20Can-Do%20Statements.pdf

Van Ek, J. A. & Trim, J. L. M. (1998). *Threshold 1990* (revised ed.). Cambridge University Press. https://lextutor.ca/vp/comp/cefr_docs/Threshold_CUP.pdf

About the Authors

Suzanne H. Rice is currently an ESL Instructor at the English Language Center at Brigham Young University, in Provo, Utah. She obtained her M.A. degree in TESOL from BYU. Her interests include written corrective feedback and TESOL pedagogy.

Grant Eckstein is a professor of linguistics at Brigham Young University specializing in TESOL. He received his Ph.D from the University of California, Davis, and his research interests include second language reading and writing development and pedagogy. He is the associate editor of Journal of Response to Writing.

Appendix A

AWL words appearing in more than one article and percentage from each sublist of the AWL. Numbers in parentheses indicate the number of articles in which each word appeared.

1	2	3	4	5	6	7	8
7.4%	6.1%	5.1%	3.3%	5.3%	3.3%	4.6%	2.6%
Area (13)	Compute (6)	Technology (9)	Investigate (5)	Challenge (7)	Reveal (10)	Globe (9)	Detect (5)
Create (9)	Tradition (6)	Physical (7)	Job (3)	Generation (6)	Expert (7)	Survive (5)	Eventual (5)
Evident (9)	Comple X(6)	Locate (6)	Predict (3)	Image (5)	Transform (5)	Decade (5)	Drama (5)
Process (8)	Region (5)	Link (5)	Code (3)	Alter (5)	Display (4)	Innovate (4)	Vehicle (3)
Percent (8)	Potential (5)	Technique (4)	Apparent (2)	Expand (5)	Transport (4)	Visible (4)	Currency (2)
Identify (8)	Conclude (5)	Layer (4)	Stress (5)	Enable (5)	Motive (3)	Confirm (4)	Intense (2)
Analyse (7)	Affect (5)	Alternative (4)	Emerge (2)	Evolve (4)	Migrate (2)	Equip (4)	Manipulate (2)
Occur (7)	Community (5)	Remove (4)	Commit (2)	Network (4)	Cooperate (2)	Chemical (3)	
Indicate (7)	Design (4)	Convene (3)	Undertake (2)	Monitor (4)	Diverse (2)	Isolate (3)	
Research (6)	Transfer (4)	Document (3)		Energy (4)		Ultimate (3)	
Issue (6)	Culture (4)	Rely (3)		Modify (3)		Release (3)	
Estimate (6)	Focus (4)	Shift (3)		Transit (2)		Media (2)	
Data (6)	Site (4)	Fund (3)		Medical (2)		Insert (2)	
Environment (5)	Consume (3)	Core (2)		Stable (2)		Couple (2)	
Individual (5)	Credit (3)	Instance (2)		Symbol (2)		Reverse (2)	
Legal (5)	Final (3)	Publish (2)		Aware (2)		Identical (2)	
Economy (4)	Strategy (3)	Comment (2)		Conflict (2)			
Structure (4)	Impact (3)	React (2)		Contact (2)			
Theory (4)	Range (3)	Sufficient (2)		Expose (2)			
Factor (4)	Invest (2)			Target (2)			
Method (4)	Aspect (2)						

9
2.3%
Team (5)
Device (3)
Military (2)
Revolution (2)
Analogy (2)
Bulk (2)

10
0.9%
Colleague (5)
Convince (3)

1	2	3	4	5	6	7	8
Source (4)	Secure (2)						
Involve (4)	Achieve (2)						
Period (4)	Construct (2)						
Require (4)	Resource (2)						
Specific (4)							
Finance (3)							
Major (3)							
Respond (3)							
Role (3)							
Function (2)							
Preceed (2)							

Exploring the Psychological Reality of Some Guidelines for Construction of Multiple-Choice Items

Sasan Baleghizadeh, Shahid Beheshti University, Tehran, Iran

Ali Momeni, Shahid Beheshti University, Tehran, Iran

Abstract

Despite the fact that multiple-choice items do not produce qualitative data and are difficult to construct, they are still widely used and one needs to follow certain guidelines to write acceptable items. However, there is some doubt about the truth of some of the guidelines to construct multiple-choice items. The present study was an attempt to investigate the truth behind three frequently-cited guidelines for the construction of multiple-choice items. To this end, sixty-nine EFL students took two versions of a 15-item vocabulary test developed by the researchers based on three well-known vocabulary item construction guidelines. The first version included well-constructed items, and the second version contained the so-called faulty items, and these labels (e.g., well-constructed, faulty) were chosen based on the target guidelines. The results of comparing the mean score of each group through an independent t-test indicated that violation of the guidelines did not result in any significant difference between the students' performance in both groups. Moreover, the lack psychological truth for the target guidelines was discovered through students' reports in the thinking aloud procedure of a verbal report.

Keywords: Testing, Vocabulary, Multiple-Choice items, Psychological reality

Introduction

Testing has always been an inseparable part of learning in general and language testing in particular. As Farhady, Jafarpoor, and Birjandy (1994) suggest, tests are used to make decisions that influence people's lives, therefore they must provide as accurate information as possible to enable learners to make fair decisions. As a result, tests are the most common evaluating process in any language teaching environment. So teachers have to equip themselves with effective test-designing strategies and techniques to improve their skills to develop language tests. Moreover, many researchers' focus has been on identifying features that affect test takers' performances. For example, many research studies (Bachman, & Palmer, 1981, 1982; Caldwell, & Pate, 2013; Shohamy, 1984) have indicated that the methods we utilize to measure language ability can affect language tests' performance. According to Bachman (1990, p. 113) "performance on language tests varies as a function both of an individual's language ability and of the characteristics of the test method."

One of the techniques which could influence the students' performance in language tests is multiple-choice items. Multiple-choice items are still one of the most widely used items in standardized testing programs and assessments done by teachers in the classroom for different reasons (Fuhrman, 1996). Every form of assessment has its advantages and disadvantages; many of the advantages of multiple-choice items have a sound basis in psychometrics, the psychological theory of mental measurement. In the meantime, teachers need to be aware of the advice and recommendations on how to construct different types of test items (see, for example, Hughes, 1989; Shultz, & Whitney, 2005; Valette, 1977; Weir, 1988). With regard to multiple-choice items, there are some guidelines on how to avoid common mistakes and pitfalls in constructing these items. The present study intends to focus on three frequently-cited guidelines for developing multiple-choice items.

The first guideline is that the test writer should not start the stem of the items with a blank (Farhady et al.1994), and this guideline could be attributed to the lack of context in the stem, which poses a problem in the item (Alderson, Clapham, & Wall 1995).

Another guideline is to avoid using alternatives that are different and stand out from other distracters (Alderson et al. 1995; Heaton 1988).

Further instruction for developing multiple-choice items is to avoid utilizing two antonyms in the alternatives of the item (Heaton 1988).

For many years, masters of language testing interested in multiple-choice items (e.g., Alderson et al. 1995; Farhady et al. 1994; Heaton 1988; Hughes 2003) have provided teachers with invaluable guidelines on writing well-constructed multiple-choice items. However, many of them have not been subject to empirical evaluation. The present study aimed to investigate the truth behind some of these guidelines and see if they would affect test-takers' performance if violated.

Literature Review

Testing has always been an undeniable force in education. In fact, students' scores on a test are used to quantify and evaluate students' achievement, teacher performance, and school progress. Sometimes it happens that these test scores under-represent what the student knows and his or her knowledge because of some reasons like student's confusion by the format, vague terminology, and inappropriate ways of asking items (Kilian, 1992).

For a long time and even today, second language teachers have been using standardized tests in language programs. As Isavi (2012, p.2) rightly points out, 'the history of second or foreign language programs have been characterized by a long tradition of standardized testing as the most reliable procedure to uncover learners' language abilities'. One of the widely used formats of items utilized in these standardized testing procedures is multiple-choice items that can be considered a selected-response test. Brown and Hudson (1999) consider two disadvantages for selected-response items, including the difficulty of constructing these types of items and the fact that they do not require students to utilize any productive language. Along the same line, Heaton (1988) states that the most significant criticism of multiple-choice items is that frequently it does not lend itself to the testing of language communication. Nevertheless, selected-response-items, especially multiple-choice items, are advantageous in measuring students' ability to recognize correct grammatical forms (Heaton, 1988). They are easy to score and more economical compared to other forms of testing (Hughes, 2003).

In the same research strand, Fuhrman (1996) asserted that prominent features such as higher reliability, lower cost of scoring, and extensive sampling of the content, are the advantages of multiple-choice items over most free-response testing procedures. Fuhrman (1996) contended that the requirement for developing well-constructed multiple-choice items has been mentioned as writing unambiguous items, giving clear directions, providing simple visual stimuli or syntax, not giving unintended clues, utilizing attainable vocabulary, and last but not least, testing concepts instead of vocabulary. By reviewing the existing literature, it can be seen that some sources have provided some considerations and guidelines for language test designers. Kehoe (1994) has provided some guidelines for composing the stems and options of multiple-choice items. For instance, the distracters should be comparable in complexity, length, and grammatical form to the answer, and the test writer should be cautious in using the words such as 'never,' 'always,' and 'all' in the stem of the items. The stem of the items should also provide as much information in the stem as possible.

Similarly, a short guide about writing effective test items, entitled 'is this a trick question' developed by Clay (2001), elaborated on some matters regarding test development and different sort of tests. The second part of this short guide attempts at actual test question construction, and five test item types are explained, and some guidelines for constructing multiple-choice items are mentioned. Some of these guidelines related to the stems of the multiple-choice items are eliminating irrelevant information in the stems and presenting explicit problems in the stems. Regarding the alternatives, it is pointed out that they should be mutually exclusive, plausible, and equal in terms of length. Besides, Carr (2011) elaborated on some of the most important guidelines for constructing multiple-choice items. Writing items with simple and clear language without any ambiguity, avoiding negatives in the items as much as possible (Haladyna, Downing, & Rodriguez 2002; Thorndike and Hagen 1969), making the alternatives distinct from each other (Cangelosi, 1990), making each distracter to be attractive to some students, and making sure that all options come from the same semantic domain, and of the similar length, were most important guidelines which were remarked by Carr (2011).

The literature reviewed above demonstrates a plethora of sources that mentioned some guidelines and considerations for developing multiple-choice items. Still, none of these guidelines is based on experimental research studies. In this regard, it can be claimed that the author of these sources mentioned these guidelines based on their experience, knowledge, wisdom. Moreover, Haladyna, Downing, and Rodriguez (2002) validated some of these guidelines through textbook authors' collective perspectives and empirical research procedures. To this end, they reviewed what was discovered in 27 textbooks on educational testing, as well as the findings of 27 research articles and reviews released since 1990. The results of their studies revealed four guidelines that received the most attention, including the use of negative words in the stem, the desirability of the number of options, use carefully none of the above, and avoid all of the above. It is of great significance to know that empirical studies are not sufficient for checking out the accuracy and correctness of the mentioned guidelines, and an experimental research study is demanded in this regard. The lack of research in this area is a remarkable gap in the literature. To address this gap, the present study is an attempt to provide an answer to the following research question:

What is the psychological reality of utilizing each (1-2-3 which appear below) multiple-choice item writing guidelines?

The justification for selecting these three guidelines lies in the fact that they are three of the most widely used guidelines that have been reported by many scholars and researchers (see Alderson, et al. 1995; Farhady et al. 1994; Heaton 1988). Furthermore, drawing on the researchers' experience participating in various examinations, these three guidelines are utilized for developing vocabulary items.

Methodology

Participants

The participants for the main study were 69 students in two classes, including males and females (45 female, 23 male) studying English language literature. Group C (control group) consisted of 34 students and group F (experimental group) 35 students; their native language was Persian. Most of them had generally spent some time learning English in their university, who were all at the same language proficiency level.

Instruments

Due to the study's aim, which was to check the accuracy and correctness of the guidelines for multiple-choice items, the researchers had to construct the tests themselves. The pre-test consisted of 15 multiple-choice items. Having used a teacher-made test, the researchers had to estimate its reliability and validity.

The reliability of the test, using the Cronbach alpha formula, was calculated to be 0.82. Some other teachers were asked to express their opinion on the content validity of the tests.

The Target Guidelines Used in the Study

For this study, the researchers chose three guidelines for constructing multiple-choice items. These three guidelines were:

1. The test writer should not start the stem of the items with a blank (Farhady et al. 1994)
2. The writer of the test should also avoid using alternatives that are different and stand out from other distracters (Alderson, et al. 1995; Heaton 1988)
3. In writing multiple-choice items the designer should not utilize two antonyms in the alternatives of the item (Heaton 1988),

Note should be taken that most of these guidelines are related to the construction of vocabulary multiple-choice items.

Procedure

Two vocabulary tests, each one including fifteen items, were developed by the researchers. The first version was the well-constructed version of the items, whereas the second one contained the poorly-constructed version of guidelines, and for each guideline, five items were designed. The two versions of the test were given to two groups (C, control group, and F, experimental group), and after administering the test and obtaining the score, the mean score of each group was compared to the mean score of another group through an independent sample t-test. Moreover, a verbal protocol procedure was carried out, to determine the mental operations responsible for a student's observed performance on a test (Ward, 2014). A useful way of studying item writing arises from cognitive psychology and entails interviewing students while they are taking a test (Norris, 1990). Moreover, the think-aloud procedure can create a situation which provides information for the cognitive processes underlying a student's encounter with a test item, and it also gets participants to verbalize their thought (Ericsson & Simon, 1993). The thinking aloud procedure was used in this study as participants were answering the items, and the researchers asked them some questions.

Participants' talks were recorded and written down. Then, the researchers analyzed the recorded protocols. Enough time was given to students to finish the test, which was measured. The average time for both groups to answer the test was twenty minutes.

Data Analysis

After administering the test to both groups and obtaining their scores on the test, to understand whether there was a statistically significant difference between the mean score of each group, an independent sample t-test was calculated through the SPSS program with the alpha level set at .05.

Results

Tables 1 and 2 illustrate the results of the t-test for the fifteen items.

Table 1: Group Statistics for Two Groups

Group Statistics					
	Groups	**N**	**Mean**	**Std. Deviation**	**Std. Error Mean**
Group 1	Group C	34	3.1714	1.36092	.23004
	Group F	35	3.0857	1.29186	.21836
Group 2	Group C	34	3.6286	1.49678	.25300
	Group F	35	4.0286	1.12422	.19003
Group 3	Group C	34	4.0571	1.05560	.17843
	Group F	35	3.9143	1.42192	.24035

Group 1 for items 1 to 5, group 2 for items 6 to 10, and group 3 for 11 to 15. Group C refers to those who received the correct version of items, whereas group F refers to those who received the faulty items.

Table 2: The Results of the T-test for both Groups

Independent Samples Test										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Total	Equal variances assumed	.345	.559	-.214	68	.831	-.17143	.80087	-1.76954	1.42668
	Equal variances not assumed			-.214	67.788	.831	-.17143	.80087	-1.76963	1.42677
Total	Equal variances assumed	.345	.559	-.214	68	.831	-.17143	.80087	-1.76954	1.42668
Group 1	Equal variances assumed	.173	.679	.270	68	.788	.08571	.31718	-.54720	.71863
	Equal variances not assumed			.270	678.16	.788	.08571	.31718	-.54723	.71866
Group 2	Equal variances assumed	4.514	.037	-1.264	68	.210	-.40000	.31642	-1.03140	.23140
	Equal variances not assumed			-1.264	63.100	.211	-.40000	.31642	-1.03229	.23229
Group 3	Equal variances assumed	2.421	.124	.477	68	.635	.14286	.29934	-.45446	.74018
	Equal variances not assumed			.477	627.45	.635	.14286	.29934	-.45537	.74109

The fifteen items were subjected to three independent samples t-test. Therefore, one independent samples t-test was run for items 1 to 5 (e.g., group 1), 5 to 10 (e.g., group 2), and 10 to 15 (e.g., group 2). Three t-tests were carried out to compare the means of students' scores for each guideline in the group that received the faulty version and the group that received the well-constructed version. The results of the independent samples t-test revealed that the significance level for all three groups exceeded the standard error of 0.05, implying that there was no significant difference between the three groups (Table 2, t (68) = 0.270, p =.788 for Group 1; t (68) = -1.26, P =.210 for Group 2; and t (68) = 0.477, p =.635 for Group 3).

The verbal protocol was conducted with the group that received the faulty version of the items. In this case, eight students were asked some questions regarding the items while they were taking the test. Two of these participants were male, and the other six participants were female. They were all at the same level of proficiency.

Guideline Number One

For guideline number one, half (4) of the students said that it has been difficult for them when the item starts with the blank. Some of them mentioned that the test designer had not provided enough information before the blank to guide them in choosing the correct response. This can be understood from the following conversation between the researchers and the students during the thinking aloud procedure:

Researchers: Dear friend, as you can see, the following item starts with a blank; did you have any difficulty answering the item? If yes or not, why?

4. ………………. to succeed is the first element of success in the life of any hard work person.

1. Confirmation
2. Determination
3. Impression
4. Permission

Student 1: It was somehow difficult for me since, in the first part, I was presented with a blank with no information, and it was somehow strange. For answering this item, I read the stem twice to be able to choose my answer.

Moreover, half of the other students (4) stated that the blank at the first part of the sentence doesn't pose any problem for them to choose the answer, and this can be understood from the below answer of one of the students about an item which started with a blank.

Student 2: It was easy for me to answer the item since I knew the meaning of the stem and all alternatives, which could help me choose the correct answer. Moreover, the existence of some clues and hints makes it possible for me to select the answer.

In this way, it might be suggested that this guideline is valid to 50 percent to apply in the test construction procedures of classroom tests.

Guideline Number Two

For this guideline, most of the students (6) reported that the existence of the different options in terms of length or difficulty level had not created any feeling to choose this option as a correct answer and they just treated it as a usual option, and this is clear in the following conversation.

Researchers: Dear friend, as you can see in the following item, option number four is different from other options in terms of length. Did this different option create any feeling to choose (or not) this option? If yes, what feeling?

9. The Police said that the …………….. cause of death was drowning, but further tests were needed.

1. apparent
2. temperate
3. domestic
4. oral

Student 3: No, I did not consider this possibility. It seems interesting to me when you mentioned this point because I did not pay attention to that. Whenever I want to answer the vocabulary items, the important thing for me is the meaning of the stem and alternatives. I knew the meaning of both the stem and alternatives in this item, so I simply chose the correct answer.

Moreover, concerning guideline number two, two students said they consider this point, and sometimes they provide answers to the items based on this point. The students reported that their justified reason for choosing this different option is based on chance, and since they do not know the meaning of the alternatives, they choose this different option, which might be the correct response.

Guideline Number Three

Regarding the third guideline, six students stated that they did not consider those pairs of antonyms, and they did not pay attention to these two antonyms words. As mentioned for the second guideline, their justified reason was that they knew the meaning of words in the stem and alternatives, which paves the way for them to choose the correct answer. Furthermore, the result of the thinking aloud procedure with another two students about this guideline revealed that these students considered these two pairs of antonyms because they did not know the meaning of the alternatives. They said that these antonyms were somehow the test designer's trick, and we thought one of these two antonyms might be the correct answer based on chance and our feeling. The verbal protocol procedure results for this guideline were somehow equal to guideline number two regarding students' attitudes and their reasons for considering or not considering those different and two pairs of antonyms words.

Based on the results of students' statements for guidelines two and three, it can be understood that these two guidelines are not that much accurate. The students who do not have enough knowledge make use of these guidelines as a trick to answer the items, but for good and knowledgeable students, these guidelines could not work since they answer the items based on their knowledge of the meaning of words.

Discussion

This study demonstrated that violation of the guidelines would not result in a statistically significant difference between the students' performance in both groups, which could possibly shed light on the lack of psychological reality and accuracy for the mentioned guidelines. Moreover, in this study, violation of the guidelines appeared to have no effect on item difficulty since there was no significant variation in terms of the student's performance in both groups.

These findings seem to suggest that the authors have mentioned these guidelines based on their experience, and there is little experimental evidence for recommending these guidelines to test writers. Examination of the result of the item analysis of a test can reveal information about the test's ambiguities, test items' effectiveness, and checking a partial accuracy of the guidelines that the test writer has used.

Besides, teachers who are interested in using multiple-choice items in their contexts should be cautious about using these guidelines, and it would be better for them if they ask a teacher or an expert in testing to check whether the items following the guidelines of item construction are reliable and accurate or not. Furthermore, it would be helpful for teachers to pilot their test, which includes those guidelines, and based on the result of the pilot test, decide to apply or not to apply those guidelines in their tests.

Moreover, of the specified guidelines analyzed in the verbal protocol procedure, only the first guideline was mentioned by some students to have an unpleasant effect on their performance. Their justified reason was acceptable and scientific, but for the other two guidelines, the students who paid attention to those guidelines were justifying their reason through chance, which is not acceptable.

The idea that the students' performance will be affected by the violated version of the guidelines for multiple-choice items is a myth; what is true is that such guidelines do not have enough reality and accuracy to influence students' performance. This might reveal the experience-based factor of these guidelines. In this regard, test designers, content, and material developers should be more attentive in providing guidelines for multiple-choice items.

Considering the ever-growing landscape of language testing, the findings of this study can be used to inform teacher educators about the necessity of training teachers so that they can be familiar with the reality behind the guidelines for multiple-choice items.

Moreover, this study's findings can help education policymakers and course designers become aware of the reciprocal relation of English learning and the effective use of techniques to examine students' abilities in testing situations.

Conclusion

The results of this study provide sufficient evidence that violations of multiple-choice item-writing guidelines resulted in no significant difference between the students' performance on the test. Additionally, these violations did not appear to influence students' performance based on the students' justified reason during the thinking aloud procedure. Although this study elucidates the effects of item construction choices, the researchers would highly recommend that while consideration and enough instruction should be given to the writers and authors of the books and researchers, these guidelines should be regarded as recommendations that can be implemented or rejected based on the test designer's discretion, rather than rules that must be enforced.

This study's findings can be of potential help and use for teachers and caregivers to better understand the use of multiple-choice items to assess students' performance. It seems increasingly important for language teachers to develop an understanding of different techniques for designing items.

Regarding the sample size of this study, it should be noted that further research with a larger number of participants at different levels would be beneficial to provide more evidence for the reality and accuracy of these guidelines.

Moreover, this study just focused on three guidelines for constructing multiple-choice items. In future studies replicating the reality and accuracy of guidelines of multiple-choice items, studies with participants with consideration of various guidelines of multiple-choice items can provide a more comprehensive and clearer picture of the reality and accuracy of these guidelines. In this way, findings can enjoy more generalizability power too.

References

Alderson, J. C., Clapham, C., & Wall, D. (1995). Language test construction and evaluation. Ernst Klett Sprachen. Cambridge: Cambridge University Press

Bachman, L. F. (1990). Fundamental considerations in language testing. Oxford: Oxford University Press.

Bachman, L. F., & Palmer, A. S. (1981). The construct validation of the FSI oral interview. Language Learning, 31(1), 67-86.

Bachman, L. F., & Palmer, A. S. (1982). The construct validation of some components of communicative proficiency. TESOL Quarterly, 16(4), 449-465.

Brown, J. D., & Hudson, T. (1999). The alternatives in language assessment. TESOL Quarterly, 32(4), 653-675.

Cangelosi, J. S. (1990). Designing tests for evaluating student achievement. White Plains, NY: Longman

Carr, N. T. (2011). Designing and analyzing language tests. Oxford: Oxford University Press.

Caldwell, D. J., & Pate, A. N. (2013). Effects of question formats on student and item performance. American Journal of Pharmaceutical Education, 77(4), 1-5. doi: 10.5688/ajpe77471

Clay, B. (2001). A short guide to writing effective test questions. Lawrence: Kansas Curriculum Center, University of Kansas.

Ericsson, K. A., & Simon, H. A. (1993). Protocol analysis. Verbal reports as data (Revised edition). MITP, Cambridge, MA.

Fahady, H., Jafarpour, A., & Birjandi, P. (1994). Testing language skills: From theory to practice. Tehran: SAMT.

Fuhrman, M. (1996). Developing good multiple-choice tests and test questions. Journal of Geoscience Education, 44(4), 379-384.

Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. Applied Measurement in Education, 15(3), 309-333.

Heaton, J. B. (1988). Writing English language tests. London: Longman.

Hughes, A. (1989). Testing for language teachers. Cambridge: Cambridge University Press.

Isavi, E. (2012). The effect of dynamic assessment on Iranian L2 writing performance. International Journal of the Learning Sciences, 40(2), 200-211.

Kehoe, J. (1994). Writing multiple-choice test items. Practical Assessment, Research, and Evaluation, 4(1), 1-3.

Kilian, L. J. (1992). A school district perspective on appropriate test-preparation practices: Areaction to Popham's proposals. Educational Measurement: Issues and Practice, 11(4), 13-26.

Norris, S. P. (1990). Effect of eliciting verbal reports of thinking on critical thinking test performance. Journal of Educational Measurement, 27(1), 41-58.

Shohamy, E. (1984). Does the testing method make a difference? The case of reading comprehension. Language Testing, 1(2), 147-170.

Shultz, K. S., & Whitney, J. W. (2005). Measurement theory in action. London: Sage Publications.

Thorndike, R. L., & E. Hagen. (1969). Measurement and evaluation in psychology and education (3rd ed.). New York: John Wiley and Sons.

Valette, R. M. (1977). Modern language testing. New Work: Harcourt Brace Jovanovich

Weir, C. J. (1988). Communicative language testing. University of Exeter.

Ward, P. (2014). Verbal protocols. In R. Eklund & G. Tenenbaum (Eds.), Encyclopedia of sport and exercise psychology (pp. 778-780). Sage Publications, Inc., https://www.doi.org/10.4135/9781483332222.n319

About the Authors

Sasan Baleghizadeh is Associate Professor of TEFL at Shahid Beheshti University in Tehran, Iran, where he teaches courses in applied linguistics, syllabus design, and materials development. He is interested in investigating the role of interaction in English language teaching and issues related to materials development. His published articles appear in both national and international journals including TESL Reporter, ELT Journal, Language Learning Journal, and Language Teaching Research.

Ali Momeni holds an MA degree in TEFL from Shahid Beheshti University in Tehran, Iran. He is interested in constructing English language examinations for EFL university students and his research areas include language assessment and CALL.

Appendices

Appendix (A): The correct and false version of the items for each guideline.

Sample items for guideline number one:

Correct version

1. According to some scientists, trying to ………… fatty foods from your diet will put you in a better state of health.
   1. eliminate
   2. utilize
   3. assemble
   4. reform

Faulty version

1. ………….. fatty foods from your diet will put you in a better state of health.
   1. Eliminating
   2. Utilizing
   3. Assembling
   4. Reforming

Sample items for guideline number two:

Correct version

6. I hope your father's advice will ……………. you to change your mind about buying that old car.
   1. condemn
   2. convey
   3. convince
   4. confess

Faulty version

6. I hope your father's advice will ……………. you to change your mind about buying that old car.
   1. condemn
   2. entrap

3. convince
4. confess

Sample items for guideline number three:

Correct version

11. There is no need to …………….. yourself about this matter; we are dealing with it.
    1. concern
    2. present
    3. suggest
    4. include

Faulty version

11. There is no need to …………….. yourself with this matter; we are dealing with it.
    1. concern
    2. present
    3. accept
    4. reject

Improving Student Evaluations of TESOL Practitioners

K. James Hartshorn, Brigham Young University, Provo, Utah, USA
Judson M. Hart, Emmersion, Lehi, Utah, USA
Benjamin L. McMurry, Brigham Young University, Provo, Utah, USA

Abstract

Teacher evaluation in second language education continues to become more important to stakeholders and is increasingly associated with higher-stakes decisions that impact teachers in substantial ways. Nevertheless, many program administrators struggle to know what to include in teacher evaluations and how best to use the results. We know very little about the kinds of factors that lead toward more favorable teacher evaluations. Thus, in addition to identifying best practices for developing and utilizing teacher evaluation instruments and data, this study sought to identify factors that lead toward better teacher evaluations. This study analyzed 5461 student evaluations of their teachers using exploratory and confirmatory regression analysis to identify eight factors with the greatest influence on student recommendation of their TESOL practitioners. Results included pedagogical practices accounting for more than half of the variability associated with teacher recommendation (adjusted $R^2 > .55$, $p < .001$). These included exemplifying professionalism, relying on course outcomes, cultivating a positive learning environment, evaluating learning effectively, optimizing class time, planning lessons effectively, utilizing homework strategically, and providing meaningful and timely feedback. Implications and applications for these findings are discussed.

Keywords: Principled pedagogical practices for TESOL, student evaluation of teachers, TESOL practitioner improvement, net promoter scores

Introduction

The importance of teacher evaluation continues to increase in many global contexts including those associated with teaching English to speakers of other languages (TESOL). Boraie (2014) suggests "Teacher evaluation is here to stay, and the question to be considered is not whether there should be a teacher evaluation system but how to evaluate teachers effectively." Moreover, evidence suggests that the stakes associated with teacher evaluation may be higher than ever before as more institutions use evaluations to shape institutional policy, practice, and employment decisions (e.g., Boraie, 2014; Fantini, 2018; Howard & Donaghue, 2015; Thomsen, 2014; Rucinski & Diersing, 2014). Efforts to improve the quality and efficacy of teacher evaluation has generated a great deal of scholarship over recent decades (e.g., Berk, 2005; Howard & Donaghue, 2015; Paulsen, 2002; Pennington & Young, 1989; Spencer & Schmelkin, 2002). What may be less clear, however, is an appropriate understanding of what teacher evaluations should include and what factors may influence student evaluations of teachers. As work continues to improve the value and utility of teacher evaluations, efforts to extract greater understanding about language learner needs evident in this medium should also increase. Therefore, this study examines thousands of recent teacher evaluations in a TESOL context to identify factors associated with teacher preparation and practice as well as the learning context that may help clarify important ESL learner perceptions. The resulting insights should be highly relevant for program administrators seeking to improve teacher evaluation, TESOL practitioners who hope to receive more favorable ratings from their students, and researchers interested in factors affecting classroom learning.

Literature Review

Relevance of Evaluation in TESOL

Teacher evaluation has evolved a great deal over the past century, impacting many TESOL contexts worldwide. Although some issues are specific to TESOL, others are applicable to much broader educational contexts. For example, early in the twentieth century in the United States, teacher evaluation generally emphasized traits such as the 83 attributes highlighted by Charters and Waples (1929). These included attractiveness, cheerfulness, dignity, health, intelligence, and so on. By mid-century, however, the focus had shifted toward better professional development as demonstrated through competency exams and certifications. Nevertheless, such credentials did not always translate to better student performance. Subsequently, reports in the United States such as A Nation at Risk (National

Commission on Excellence in Education, 1983) precipitated a shift in thinking toward learning outcomes. This was further solidified by the No Child Left Behind Act (NCLB, 2002), which marshalled in a new era of standardized testing and evaluation. The same influences that precipitated increased standardized testing in the United States impacted other countries as well, and by 2006, the number of nations incorporating standardized testing more than doubled compared to the decade before (Benavot & Tanner, 2007).

This new global culture of educational assessment and evaluation continues to impact teachers in many ways. For example, since 2009 nearly 65% of the states in the United States have dramatically overhauled their systems for teacher evaluation, which have had a substantial impact on institutional policy (e.g., Thomsen, 2014) as well as how institutions hire, fire, train, assess, and reward their teachers (Rucinski & Diersing, 2014). Nevertheless, the debate over whether these changes actually help students achieve learning outcomes remains contentious. There are numerous kinds of ESL/EFL programs around the world in a wide variety of settings ranging from government-sponsored schools for children or adults to community programs designed to help residents or refugees develop the English language skills they need to gain employment and flourish within in society. This vast diversity of contexts in which these programs operate may make the processes and results of this work particularly valuable. Regardless of whether the context may be ESL or EFL, whether working with children or adults, or whether educators are teachers or administrators, language educators are increasingly more likely to grapple with a variety of issues associated with teacher evaluations.

The growing importance and use of teacher evaluations in many language education contexts raise many crucial questions such as What should evaluations include? How can administrators ensure that they are fair? What should be done with evaluation data once it is collected? While many institutions continue to wrestle with these kinds of questions, we add an additional question of particular relevance to this study, What insights can evaluation-based research provide to the field of language teaching? We believe that a systematic study of carefully designed teacher evaluations may unlock valuable insights regarding student experience that could impact language development.

Effective Teacher Evaluation

In considering the benefits of evaluation-based research, it is important to remember that attempts to measure the contribution of a teacher on student learning will be complex and multidimensional. For example, some researchers estimate that the actual influence a teacher has on student performance ranges from 14% down to 1% depending on a wide array of factors (American Statistical Association, 2014). Therefore, we side with scholars such as Mathis (2012) and Boraie (2014) who suggest that effective teacher evaluation needs to be established on a multitude of factors. Although it may be appropriate to examine actual student achievement of learning outcomes as part of a broader evaluation, this should be done in view of the specific teaching and learning context, recognizing that even exceptional teachers may have little control over student performance. A broader body of evaluation might also include observations conducted by administrators and peers, citizenship measures associated with dependability, collegiality, productivity, mentoring, and cooperation as well as the learners' evaluations of their teachers. While these and many other factors may be important to examine as part of a broader teacher evaluation strategy, this study focuses exclusively on the tangible instrument used by students to evaluate their teachers.

We acknowledge that some teachers are skeptical regarding the validity and utility of evaluations generated by students. Concerns may range from assumptions that evaluation results are simply a function of teacher popularity or that easy classes will produce higher evaluation scores. Although students may lack expertise in a number of areas relative to overall teacher performance, the students themselves may be in the strongest possible position to evaluate the effectiveness of many aspects of their classroom experience. Moreover, many scholars such as Aleamoni (1999) have debunked numerous myths about student evaluations of teachers, providing strong evidence for the reliability and validity of "well-constructed instruments and procedures" (p. 155) and that such evaluation results can be used effectively to improve teaching.

With the intent of improving our own teacher evaluations and procedures associated with our intensive English program, we analyzed our instruments, carefully weighing relevant literature and the specific needs of our own institution. Our approach was consistent with recommendations of Spooren, Brockx, and Mortelmans (2013) who suggested that when developing evaluation instruments, “institutions should be able to select the aspects that are most important, according to their educational vision and policy” and that many stakeholders such as “administrators, teachers, and students…should be involved in the definition of these characteristics” (p. 603). Through this process, we concluded that our teacher evaluations should include two elements that were not currently in use but that were needed to optimize the effectiveness and utility of the instrument. The first was an overall summary measure of performance that could be used across a variety of teachers and contexts. The second was to replace more generic indicators of performance with very focused areas of specific interest to the institution. The relevance and applications of these two components are described below.

Summary score of student sentiment

One challenge with many evaluations is the lack of an effective indication of overall student sentiment. While perhaps incomplete, a general sense of practitioner performance may help simplify evaluation and provide administrators with broad insight. Research suggests that a single overall score for teacher evaluation can be well correlated with other important measures and may be useful for summative purposes (Cashin & Downey, 1992; Fantini, 2018). An overall score facilitates broad comparisons across teachers and courses over time, allowing administrators to identify relevant trends and to make appropriate programmatic adjustments.

In this regard, the world of business may have a useful metric. Today numerous companies ranging from airlines and retail to healthcare and technology use what is called a net promoter score (NPS) to provide an overall evaluative indication of the products or services they provide (Reichheld, 2003, 2006). Customers are presented with a single question to answer using a scale of zero to ten: “How likely is it that you would recommend [company, product, or service] to a friend or colleague?” Proponents describe the NPS as providing essential insights (e.g., Martin, 2011; Reichheld & Markey, 2011), and that it has been used with great success across a wide array of businesses such as banking, cosmetology, telecommunications, healthcare, and so on (e.g., Hamilton, Lane, Gaston, Patton, MacDonald, Simpson, & Howie, 2014; Leisen Pollack & Alexandrov, 2013).

Traditionally, respondents providing a nine or ten were labeled promoters, those responding with a seven or eight were considered passives, and those responding with a one through six were identified as detractors. Typically, the NPS has been calculated by subtracting the proportion of promoters from the proportion of detractors while ignoring the passives (e.g., 70% promoters, 20% passives, and 10% detractors would yield an NPS of 60). Though its lack of sophistication has been seen by some as a considerable limitation (e.g., Krol, Boer, Delnoij, & Rademakers, 2015; Mandal, 2014), its simplicity has also been seen as its greatest strength by many proponents and critics alike (Bendle & Bagga, 2016). Many managers like the ease with which it can be elicited by clients and then analyzed and interpreted across a variety of contexts. Though widely popular in corporate America today, the NPS is not without controversy.

Because the 11-point scale is collapsed into just three categories and the passive scores are ignored, critics have noted that important numerical information is lost when the score is calculated (e.g., Bendle & Bagga, 2016; Schneider, Berent, Thomas, & Krosnick, 2008). They suggest that utilizing the complete scale may provide more evaluative insight. Thus, in an attempt to leverage the practical benefits of an overall score in our own teacher evaluation, we determined to calculate the average of the entire 0-10 scale, based on the same question regarding the respondent's likelihood of recommending the teacher to a friend or colleague. We also determined to strengthen the evaluation by connecting the score to additional quantitative and qualitative information provided by the respondents. We refer to our modified overall score based on the complete 0-10 scale as our teacher recommendation score.

Identifying Core Pedagogical Practices

In addition to the use of an overall summary score based on student recommendation of the TESOL practitioner, the second improvement to our student evaluation was the inclusion of specific pedagogical practices addressed in the literature and consistent with our values, culture, and beliefs about language learning. Before teachers can be evaluated effectively, an institution must have a clear idea of the criteria they would like to use. Though some programs are forced to use generic evaluations, this may not be the best approach (e.g., Hill & Grossman, 2013; Hunt, Gurvitch, & Lund, 2016; Pratt, 2002). Rather than using general standards that may be unrelated to core ideals for effective practice in specific contexts, standards must be clearly articulated and should grow out of the underlying purpose for the teaching and learning within a particular program setting. They should be consistent with the collective values and beliefs of the institution and be rooted in insights gleaned from theory and practice. There will also need to be an appropriate way to measure performance levels for each of these standards.

With these ideals in mind, we set out to identify key pedagogical practices that could be used for training and evaluation purposes specific to our TESOL context. We sought to identify the core pedagogical practices that best capture and reflect our institutional values, beliefs, and aspirations for teacher performance. The process drew on input from administrators, teachers, staff, and other stakeholders. These practices were refined in collaborative meetings over the course of many months and were based on the practical experience of the stakeholders as well as relevant literature. In working through this process, an attempt was made to balance the desire for a comprehensive list of the most consequential practices in second language teaching and learning with the need to keep the set of points simple enough to ensure it could be easily conceptualized and successfully implemented.

The final list included eight pedagogical practices designed to guide classroom preparation and teaching. Though the scientific evidence of the benefits of these practices may be stronger for some than for others, the consensus of stakeholders was that each practice that survived the winnowing process was very important to our context. An abundance of literature also made a strong case for the relevance of each. The eight pedagogical practices are listed here with an abbreviated sample of relevant literature (See Appendix A for more detail regarding each practice).

1. Rely on course outcomes (e.g., Basturkmen, 2010; Leung, 2012; Richards, 2013).
2. Plan lessons effectively (e.g., Baecher, Farnsworth, & Ediger, 2014; Liyanage, & Bartlett, 2010; Milkova, 2012; Pang, 2016).
3. Optimize class time (e.g., Calderón, Slavin, &, Sánchez, 2011; Murray & Christison, 2010; Rossiter, Derwing, Manimtim, &, Thomson, 2010; Tan, Nabb, Aagard, &, Kim, 2010; Walsh, 2006).

4. Cultivate a positive learning environment (e.g., Brown, 2006; Oxford, 1999; Tsiplakides, & Keramida, 2010; Tsui, 1996; Young, 1991).
5. Evaluate learning effectively (e.g., Abedi, 2010; Bailey & Heritage, 2014; Clark, 2012; Frey, Schmitt, &, Allen, 2012; Ketabi & Ketabi, 2014; McMillan, 2013).
6. Utilize homework strategically (e.g., Gershenson & Holt, 2015; Maltese, Tai, &, Fan, 2012; McReynolds, 2010; Walberg, Paschal, & Weinstein, 1985; Wallinger, 2008).
7. Provide meaningful and timely feedback (e.g., Hartshorn & Evans, 2015; Fordham, 2015; Su & Tian, 2016).
8. Exemplify professionalism (e.g., Alsalahi, 2015; Farrell, 2015; Lorimer, & Schulte, 2012; Orlich, Harder, Trevisan, Brown, &, Miller, 2016; Sawyer, Andzik, Kranak, Willke, Curiel, Hensley, & Neef, 2017; Vu, 2016).

Once these practices were established, they were distributed to teachers, posted in hallways, emphasized in teacher training and in-service meetings, and implemented as the focus of teacher observation and professional development activities. Thus, rather than utilizing a teacher evaluation that was detached from daily practice, our updated instrument was designed to reflect practices at the very core of what we expected of our teachers on a daily basis.

Although this list of pedagogical practices was specifically designed for our unique setting, these practices should be highly relevant for most language teaching and learning contexts. Nevertheless, we recognize that some institutions might benefit from placing greater emphasis on particular practices identified here or from focusing on other components that may not have been included in our list.

Development of the Evaluation Instrument

The new teacher evaluation instrument made use of several item types to extract both quantitative and open-ended data from students. A question for each pedagogical practice was framed according to the frequency with which a student felt that the teacher effectively demonstrated that practice. Since the pedagogical practices were originally articulated for teachers and administrators rather than for students, some language was simplified in the instrument to help ensure student understanding (see Appendix B). For example, for optimizing class time, students responded to a statement such as, “My Reading teacher, Ms. Jones, gets the most from each class hour.” The student would then affirm the frequency by selecting from choices: always (5), almost always (4), usually (3), sometimes (2), or rarely (1). Depending on the student’s selection of frequency a follow-up open response question appeared in the electronic form. Thus, if a student selected always, the survey then asked for the student to Provide an example of how Ms. Jones gets the most from each class hour. If students selected sometimes, they were asked to provide a suggestion how they feel Ms. Jones could get more from each class hour in the future. This process of discrete item followed by open-response item was repeated for each of the pedagogical practices. Students were required to respond to the discrete items. They were not required to supply evidence to support their rating of each practice but were given the opportunity to do so.

Following questions about teacher application of the pedagogical practices, the student provided an overall score of teacher performance, “On a scale of 0-10 how strongly would you recommend, the ELC use Ms. Jones to teach a class similar to this reading class again?” With the summary score and pedagogical practices incorporated into the evaluation instrument, we can pursue important questions that may benefit the broader field. Although the pedagogical practices identified by stakeholders were seen as important, it was unclear how the students perceived the relative importance of these practices within their own learning experience. Nor was it clear what additional factors beyond the pedagogical practices might impact teacher recommendation scores.

These additional factors included demographic variables related to the background of the teachers or students. For example, literature suggests that teacher practice might be influenced by the level of teacher preparation (e.g., Hartshorn, Evans, & Tuioti, 2014). Moreover, researchers have observed students who express concerns that their language teacher is not a native speaker of their target language (e.g., Braine, 2013). In addition, it is conceivable that other student factors that may impact language learning could also impact teacher recommendation such as student age, gender, proficiency level, and so on (e.g., DeKeyser, 2013; Dörnyei, 2014). Answering such questions could help address a significant gap in the literature and provide important insights about student perspectives of pedagogical practices as well as general insights regarding student recommendations of their TESOL practitioners.

Research Question

With this review of literature in place, we now consider the specific research question articulated for this study: How well do student perceptions of teacher performance within the eight pedagogical practices and other demographic and programmatic factors account for teacher recommendation scores?

Methods

This section briefly addresses the data elicitation, the students who completed evaluations, and the analyses used in this study.

Data Elicitation

In order to address the research question, data from 5461 student evaluations were collected electronically by the institution examined in this study. This included sending the evaluations out to student email addresses through Qualtrics software one week prior to the end of the semester. Students in the Novice-High and Intermediate-Low Levels completed the evaluations in the computer lab to ensure additional language support if needed. Those who did not complete the evaluations right away were given reminders. The completion rate was over 98%. Any identifying student information was stripped prior to data being made available for analysis. The institution's Internal Review Board authorized the use of these preexisting data for study.

Students

Data used in this study was produced by students with a mean age of 25 (SD=7.31). Males made up 42% of the students while females made up 58%. Of the evaluations gathered for this study, 54% were from students continuing their study at the institution after at least one semester of previous study, and 46% were from new students who had just completed their first semester in the program. Evaluations came from students with various L1 backgrounds including Spanish (61%), Portuguese (10%), Chinese (9%), Korean (8%), Japanese (4%), Russian (2%), Mongolian (1%), French (1%), Thai (1%), and other languages represented with less frequency (i.e., Albanian, Arabic, Chuvash, Farsi, Hungarian, Italian, Kazakh, Lithuanian, Loa, Malagasy, Tagalog, Tajik, Ukrainian). The language proficiency levels of the students, based on ACTFL guidelines (ACTFL, 2012), is presented in Table 1.

Table 1 Student Proficiency by Evaluation Percentages

Proficiency level	%
Novice High	1.53%
Intermediate Low	5.40%
Intermediate Mid	14.13%
Intermediate High	35.94%
Advanced Low	23.73%
Advanced Mid	16.72%
Advanced High	2.56%
Total	100.00%

Analyses

Teacher recommendation scores were based on means from the 0-10 scale described previously. Answering the research question involved multiple linear regression where the eight pedagogical practices and other relevant factors functioned as the explanatory variables for the response variable—the teacher recommendation score. Table 2 presents the list of variables that were included. They represent topics found in the literature and ideas of specific interest within the program. For convenience, these variables are organized into five different categories. The first category includes the eight pedagogical practices described previously.

Table 2 Explanatory Variables for Predicting Teacher Recommendation

Category	Explanatory Variables
Practices	1. Rely on course outcomes 2. Plan lessons effectively 3. Optimize class time 4. Cultivate a positive learning environment 5. Evaluate learning effectively 6. Utilize homework strategically 7. Provide meaningful and timely feedback 8. Exemplify professionalism
Teacher	1. Educational level of the teacher 2. Whether the teacher was full-time 3. Whether the teacher was a native speaker of English 4. Whether it was the teacher's first semester at the institution 5. Skill area taught (reading, writing, listening/speaking, grammar)
Student	1. Student age 2. Student sex 3. Student proficiency level 4. Total number of semesters at institution 5. Hours of homework completed per week 6. Whether the student was new or returning 7. Student confidence regarding skill improvement 8. Whether the student felt challenged in the course 9. Student's level of overall satisfaction with the course 10. Subsequent semester plans (i.e., stay, vacation, leave)
Exams	1. Final exam reading score 2. Final exam writing score 3. Final exam listening score 4. Final exam speaking score 5. Final exam vocabulary score 6. Combined final exam score 7. Class Proficiency grades 8. Class Citizenship grades 9. Teacher rating of student progress
Time	Class time (i.e., 8:15, 9:30, 12:15, 1:30)

The second category includes variables associated with the teacher. Level of education was based on a four-point scale according to the following: baccalaureate degree in progress (1), baccalaureate degree completed (2), master's degree in progress (3), master's degree completed (4). No additional distinctions were made for the few teachers who were pursuing or who had completed doctoral degrees.

The third category includes student demographic information such as student age, sex, language proficiency level (based on placement testing), the number of semesters the student had studied at the institution, and whether the student was new to the institution or continuing their study. It also included self-reported information such as the number of hours spent completing homework each week, how confident students were that they improved in the language skill taught in the course, the extent to which students felt challenged in the course, the students' level of satisfaction with the course, and their plans after the completion of the semester (i.e., whether they intended to continue studying at the institution, take a vacation for a semester and then return, or leave the institution altogether).

The final categories represent several measures of student performance and the time of day the respective classes were held. Measures of student performance includes institutional proficiency exams administered at the end of the semester in areas such as reading, writing, listing, speaking, and vocabulary. It also includes class proficiency grades based on student performance over the course of the semester and citizenship grades based on class participation and homework completion. The last measure of student performance was a single, overall rating provided by the teacher of student proficiency. The final category was based on the time of day the respective classes were held. Classes met Monday through Thursday for 65 minutes at 8:15 a.m., 9:30 a.m., 12:15 p.m., and 1:30 p.m.

Because of the exploratory nature of the research question, data were randomly divided into two halves. The first half (2731 evaluations) was used to run an exploratory stepwise regression analysis, and then the second half (2730 evaluations) was used to run a confirmatory regression analysis to test the model identified through the first stepwise analysis (see Mark & Goldberg, 2001). Because stepwise regression can be prone to overfitting and distorted p-values, the default criteria in the SPSS software was adjusted to help minimize distortion and ensure that any variables associated with teacher recommendation would be truly meaningful (the typical variable entry, p=.05, and removal, p=.10, was replaced with p=.001 for entry and p=.002 for removal based on a Bonferroni adjustment for the number of explanatory variables used in the study, see Wilkinson & Dallal, 1981).

Results and Discussion

Exploratory and Confirmatory Analyses

The research question addressed the extent to which the eight pedagogical practices and other demographic or programmatic factors influenced teacher recommendation scores. Table 3 presents the results of the exploratory and confirmatory regression analyses. Of the 41 variables included, the analysis generated a model of nine variables accounting for teacher recommendation. These are listed on the left side of the Table 3. Statistics associated with the exploratory analysis are included in the middle section of the table and account for more than 55% of the variability associated with teacher recommendation (adjusted R^2 of .558). The exploratory model included all eight of the pedagogical practices discussed previously along with one additional variable associated with the combination of the student final exam scores. In general, the exploratory analysis suggests that the greater the presence of these model factors, the greater the recommendation of a given teacher.

Table 3 Factors Impacting Teacher Recommendation

	Exploratory				Confirmatory			
Model	*B*	*SE*	*ß*	*p*	*B*	*SE*	*ß*	*p*
(Constant)	-1.398	0.223		<.001	-1.713	0.173		<.001
Exemplify professionalism	0.920	0.079	0.231	<.001	0.699	0.064	0.181	<.001
Rely on course outcomes	0.572	0.079	0.149	<.001	0.563	0.066	0.147	<.001
Cultivate positive environment	0.512	0.074	0.133	<.001	0.499	0.061	0.131	<.001
Evaluate learning effectively	0.326	0.074	0.094	<.001	0.536	0.057	0.160	<.001
Optimize class time	0.308	0.068	0.090	<.001	0.421	0.059	0.119	<.001

	Exploratory				Confirmatory			
Model	*B*	*SE*	*ß*	*p*	*B*	*SE*	*ß*	*p*
Plan lessons effectively	0.272	0.078	0.078	<.001	0.460	0.063	0.130	<.001
Utilize homework strategically	0.288	0.059	0.088	<.001	0.317	0.050	0.095	<.001
Provide meaningful feedback	0.264	0.055	0.087	<.001	0.162	0.047	0.052	.001
Combined final exam scores	0.138	0.028	0.068	<.001	0.100	0.023	0.048	<.001
	Adjusted R^2=.558				Adjusted R^2=.662			

Note that the standardized betas (β) in the table show the relative importance of each of the variables to the overall model. For example, exemplifying professionalism is the most important part of the model produced by the exploratory analysis followed by relying on course outcomes and cultivating a positive learning environment. Although all of the factors provide a meaningful contribution to the model, exemplifying professionalism carries about one and a half times the weight of the next most important factor, relying on course outcomes, and more than three times the weight of the combined final exam scores.

The results of the confirmatory analysis are included in the right portion of Table 3. Though based on the same variables identified in the exploratory analysis, the confirmatory analysis accounted for just over 66% of the variability associated with teacher recommendation (adjusted R2 of .662). Though the confirmatory analysis accounted for a little more variation, than the exploratory analysis, these results seem comparable and tend to underscore the importance of the model factors in teacher recommendation.

Additional insights may be gleaned by examining just a few of the responses to the follow-up questions included in the evaluation that focus on the pedagogical practices valued the most highly by the students. For example, when asked to provide evidence of professionalism, students often commented on the teachers' overall attitude toward teaching and concern for student success. One student described his teacher with representative comments such as "very professional" and "punctual," and, in addition to planning classes well, she had "a very good attitude," and was "interested in the improvement of the students." Many other learners provided evidence of professionalism by referring to other pedagogical practices included in our list. Comments like, "she's well prepared and [uses] time wisely," were frequent throughout the data. Many students also referred to the demeanor of their teachers. They generally connected professionalism with kindness, patience, and dedication among other valued attributes.

Students who provided evidence of the teacher relying on course outcomes generally emphasized one or more of the following—introducing course outcomes at the beginning of the semester, writing or stating course outcomes at the beginning of a class period, or using activities or resources that were clearly connected to the outcomes. One representative response illustrates this with the description, "At the beginning of the course, she explained the outcomes of the class and how we were supposed to reach them." The student continued by indicating that in class they always were engaged in "activities to help us to [reach] those outcomes." Another student response echoed that of many others, "She wrote [on] the board the outcomes for the day." As with professionalism, many comments were connected to other pedagogical practices such as, "She always prepares and really focuses on the course outcomes."

Other explanatory variables had no apparent influence on teacher recommendation. Clarifying that these variables did not impact recommendation may be as important to understanding teacher recommendation as the variables that did affect it. Variables not impacting teacher recommendation include level of education, full-time versus part-time status, whether it was the teacher's first semester at the institution, and the language skill taught by the teacher (i.e., reading, writing, listening/speaking, or grammar). These findings suggest that levels of education and experience were not valued by students as much as effective practice. These findings also indicate that non-native speakers were recommended just as much as native speakers, a conclusion consistent with other scholars who have recognized potential strengths in native and nonnative teachers alike, but no disadvantage for non-native teachers (e.g., Braine, 2013; Shin & Manochphinyo, 2017; Walkinshaw & Oanh, 2014).

There were many additional factors which similarly had no observable impact on teacher recommendation. Some of these were associated with the students themselves such as student age, sex, language proficiency, time spent on homework, how long the student had studied at the institution, how challenged the student felt by the course, or the students' plans for the subsequent semester. Additional areas with no impact on teacher recommendation suggest student ability to separate the evaluation of their teachers from other conspicuous elements of their classroom experience. For example, teacher recommendations were unaffected by the extent to which students felt they improved in the specific skill area taught or their overall level of satisfaction with the course. Other factors with no apparent impact on teacher recommendation included when classes were held and eight of the nine measures of student assessment. These metrics included various classroom-based evaluations as well as student performance on institutional exams for each language skill.

Although performance on individual exams assessing the discrete language skills of reading, writing, listening, speaking, and vocabulary were not associated with teacher recommendation, the combined scores for these exams were meaningful enough to be included in the model. It should be noted, however, that the contribution of the combined exam scores provided the weakest impact to the model in both the exploratory and confirmatory analyses. Since exam performance in this study was tied to specific proficiency levels, this factor of combined exam scores is essentially an overall measure of proficiency.

An important question is why there might be a relationship between higher proficiency levels and teacher recommendations, especially when anecdotal evidence suggests that the highest proficiency learners may be more demanding and have higher expectations than their lower-proficiency counterparts. Although the answer is not obvious, here are two possibilities. It could be that higher proficiency students are able to have more meaningful interactions with their teachers in ways that foster greater language development that is recognized by the students. It could also be that if higher-proficiency learners maintain higher expectations of their teachers, the institution may place some of their best teachers in those positions where they garner positive evaluation data while teachers with less developed skills may opt out of those more challenging assignments. Additional research may be needed to better understand these possibilities.

Implications

Of the numerous variables examined in this study, the eight pedagogical practices established by the observed institution accounted for more than half of the variability associated with teacher recommendations. The relative impact of these practices within the regression model is summarized in Table 4, which averages the contributions of each factor from the exploratory and confirmatory analyses into an overall set of percentages. Although some of these variables are more important to the model than others, collectively they provide meaningful insight into those teacher practices highly regarded by the ESL learners in this study. These findings should be valuable to program administrators, teachers who may be considering ways to better meet student expectations, and researchers interested in factors affecting classroom learning.

Table 4 Contribution Percentages for Factors Explaining Teacher Recommendation

Teacher recommendation factor	Percent
Exemplify professionalism	19.80%
Rely on course outcomes	14.22%
Cultivate positive environment	12.69%
Evaluate learning effectively	12.21%
Optimize class time	10.04%
Plan lessons effectively	10.00%
Utilize homework strategically	8.79%
Provide meaningful feedback	6.68%
Combined final exam scores	5.57%
Total	100.00%

Using Evaluation Data

This study presents specific findings that should provide practitioners and administrators with useful insights as they seek to identify teaching strengths and areas for improvement. Focusing on these practices may also be beneficial in teacher preparation, teacher in-service training, or other professional development activities. Although application of these specific findings may improve the efficacy of language teaching in certain contexts, other well-designed teacher evaluations are also likely to benefit teachers who seek to improve their practice. Nevertheless, the ultimate efficacy of such instruments may depend on how the results are applied. Thus, we briefly address considerations necessary for effectively utilizing results from this study as well as other kinds of teacher evaluations used in other contexts.

Practitioners and administrators should keep in mind that the appeal for students to participate in evaluation of their teachers is largely to provide feedback to improve teaching (Chen & Hoshower, 2003). Yet, the research shows that evaluation by itself does not produce more effective teaching (e.g., Ballantyne, Borthwick, & Packer 2000; Kember, Leung, & Kwan 2002; Spooren, Brockx, & Mortelmans, 2013). There may be a number of reasons why this is true. Some practitioners do not seek to improve their teaching based on evaluation results because they may not know how to do so (e.g., Arthur 2009; Bamber & Anderson 2012; Dresel & Rindermann 2011). Some teachers may simply disregard evaluation results because of concerns regarding their validity or utility (e.g., Simpson & Siguaw, 2000). Still others may see evaluations strictly as summative. They may think that conspicuous use of evaluations for improvement may highlight their weaknesses (Baber & Anderson, 2012). Finally, other teachers simply may not have the motivation to improve the quality of their teaching (Edström 2008; Hendry, Lyon, & Henderson-Smart, 2007).

If improved teaching is the ultimate purpose of the evaluation, practitioners and administrators need to understand the conditions that are necessary for evaluation data to lead toward improved teaching. We address just three of these conditions here. First, evaluation instruments must be designed and used in ways that allow them to effectively capture and present meaningful information to the stakeholders. If evaluation instruments are perceived as ineffective by stakeholders, they should be reevaluated. Where appropriate, faulty instruments should be improved or replaced. Nevertheless, if institutions have exercised due diligence in successfully creating instruments that give the students a valued voice and effectively provide teachers with relevant, constructive, and informative data, students will be more inclined to participate and teachers are likely to glean many useful insights.

Second, practitioners must have a desire to continue to develop the quality of their teaching. Nasser and Fresko (2002) observe that utilizing evaluation results to improve practice requires a certain disposition and willingness of the teacher. Golding and Adam (2016) refer to this as an "improvement attitude" (p. 5). This is in contrast to a perception where practitioners consider their teaching to be good enough. This characteristic of successful teachers who want to improve is illustrated by practitioners who are student-centered and who look to evaluation data for insights regarding adjustments they could make to better help students meet learning outcomes (McGowan & Graham, 2009; Hendry, Lyon, & Henderson-Smart, 2007). Golding and Adam (2016) noted that the highly effective teachers they observed utilized "a reflective approach to their teaching" (p. 5) where they consistently asked themselves how they could improve, and they conscientiously looked to their evaluation data as formative feedback to help them adjust their practice in ways that benefited their students.

The final notion we address here is that practitioners are more likely to improve their teaching when they meet regularly with program administrators to review their evaluation results and to develop and act on a plan to implement appropriate adjustments. Spooren, Brockx, and Mortelmans (2013) suggest that rather than being left to analyze, interpret, and apply evaluation data on their own, "teachers should be able to rely on expert consultation" (p. 628) regarding the content of their evaluations, which might include discussing results with "colleagues or educational experts" (p. 623). Lang and Kersting (2007) observed that providing teachers with evaluation data without consultation was not effective over the course of multiple semesters. On the other hand, Dresel and Rindermann (2011) noted that consultations with teachers regarding their evaluation results had a positive effect on the quality of teaching. Similarly, Penny and Coe (2004) observed that teaching efficacy improved as practitioners were given opportunities to reflect and discuss their teaching. Thus, we encourage administrators to ensure that their teacher evaluation instruments are designed well and used appropriately, that the teaching and learning environment encourages a focus on improvement, and that teachers are given ample opportunities to process and apply insights from evaluations with input and support from administrators.

Limitations and Future Research

Though this was a fairly large study which included thousands of evaluations, it took place at only one institution where more than half of the student participants were native speakers of Spanish. Thus, additional studies at other institutions and in other contexts should be pursued in an effort to test the pervasiveness of these findings. Moreover, we acknowledge the potential effects of simplifying the language of the data collection instrument and of representing each of the practices with a single survey item. Using multiple items for each of the practices would likely increase the validity and reliability of these constructs. Further study could also examine the possible effects of L1 or national background on teacher recommendations. In addition, future research could incorporate other factors that may impact teacher recommendation that were not part of this study. Finally, it could also seek to provide additional insight regarding the possible relationship between student proficiency level and teacher recommendation observed in this study.

Conclusion

This study highlights the growing importance of teacher evaluation in TESOL contexts. It suggests that evaluation instrument development should be anchored to relevant literature and practices that best capture an institution's values, beliefs, and aspirations for teacher performance, drawing on input from administrators, teachers, staff, and other stakeholders. It suggests that teacher evaluations should elicit both qualitative and quantitative data that is informative to both administrators and the teachers themselves.

This study also examined research questions about desirable pedagogical practices and their relationship to teacher recommendations. The results suggest that higher teacher recommendations may have less to do with specific elements of the program or the demographic background of students or teachers, and more to do with the extent to which teachers successfully implement specific pedagogical practices within the classroom. These include exemplifying professionalism, relying on course outcomes, cultivating a positive learning environment, evaluating learning effectively, optimizing class time, planning lessons effectively, utilizing homework strategically, and providing meaningful feedback. These insights should be leveraged by administrators and practitioners to better understand and meet the expectations and learning needs of their students.

References

Abedi, J. A. M. A. L. (2010). Research and recommendations for formative assessment with English language learners. Handbook of formative assessment, 181-197.

ACTFL. (2012). ACTFL Proficiency Guidelines 2012. White Plains, NY: American Council on the Teaching of Foreign Languages.

Aleamoni, L. M. (1999). Student rating myths versus research facts from 1924 to 1998. *Journal of Personnel Evaluation in Education, 13*, 153-166. doi.org/10.1023/A:1008168421283

Alsalahi, S. M. (2015). Stages of teacher's professionalism: how are English language teachers engaged? *Theory and Practice in language studies, 5*(4), 671. doi.org/10.17507/tpls.0504.01

American Statistical Association (2014). ASA Statement on Using Value-Added Models for Educational Assessment. April 8. www.amstat.org/policy/pdfs/ASA_VAM_Statement.pdf

Arthur, L. (2009). From performativity to professionalism: lecturers' responses to student feedback. *Teaching in Higher Education, 14*, 441-454. doi.org/10.1080/13562510903050228

Baecher, L., Farnsworth, T., & Ediger, A. (2014). The challenges of planning language objectives in content-based ESL instruction. *Language Teaching Research, 18*(1), 118-136. doi.org/10.1177/1362168813505381

Bailey, A. L., & Heritage, M. (2014). The role of language learning progressions in improved instruction and assessment of English language learners. *TESOL Quarterly, 48*(3), 480-506. doi.org/10.1002/tesq.176

Ballantyne, R., Borthwick, J., & Packer, J. (2000). Beyond student evaluation of teaching. *Assessment & Evaluation in Higher Education, 25*, 221-236. doi.org/10.1080/713611430

Bamber, V., & Anderson, S. (2012). Evaluating learning and teaching: institutional needs and individual practices. *International Journal for Academic Development, 17*, 5-18. doi.org/10.1080/1360144X.2011.586459

Basturkmen, H. (2010). *Designing Courses in English for Specific Purposes*. New York: Palgrave Macmillan. doi.org/10.1057/9780230290518

Benavot, A., & Tanner, E. (2007). The growth of national learning assessments in the world, 1995-2006. Background paper for the Education for All Global Monitoring Report 2008: Education for All by 2015: Will we make it? UNESCO.

Bendle, N. T., & Bagga, C. K. (2016). The metrics that marketers muddle. *MIT Sloan Management Review, 57*(3), 73.

Berk, R. A. (2005). Survey of 12 strategies to measure teaching effectiveness. *International journal of teaching and learning in higher education, 17*(1), 48-62.

Boraie, D. (2014). Nine Features of an Effective Teacher Evaluation System [Blog post]. Retrieved from http://blog.tesol.org/nine-features-of-an-effective-teacher-evaluation-system/

Braine, G. (Ed.). (2013). *Non-native educators in English language teaching*. Routledge. doi.org/10.4324/9781315045368

Brown, H. D. (2006). *Principles of Language Learning and Teaching* (5th ed.). New York: Pearson ESL.

Calderón, M., Slavin, R., & Sánchez, M. (2011). Effective instruction for English learners. *The Future of Children, 21*(1), 103-127. doi.org/10.1353/foc.2011.0007

Cashin, W.E., & Downey, R.G. (1992). Using global student rating items for summative evaluation. *Journal of Educational Psychology, 84*, 563-572. doi.org/10.1037/0022-0663.84.4.563

Charters, W. W. & Waples, D. (1929). The commonwealth Teacher-Training Study. Chicago: University of Chicago.

Chen, Y., & Hoshower, L. (2003). Student evaluation of teaching effectiveness: An assessment of student perception and motivation. *Assessment and Evaluation in Higher Education, 28*, 71-88. doi.org/10.1080/02602930301683

Clark, I. (2012). Formative assessment: Assessment is for self-regulated learning. *Educational Psychology Review, 24*(2), 205-249. doi.org/10.1007/s10648-011-9191-6

DeKeyser, R. M. (2013). Age effects in second language learning: Stepping stones toward better understanding. *Language Learning, 63*(s1), 52-67. doi.org/10.1111/j.1467-9922.2012.00737.x

Dörnyei, Z. (2014). *The psychology of the language learner: Individual differences in second language acquisition*. Routledge. doi.org/10.4324/9781410613349

Dresel, M., & Rindermann, H. (2011). Counseling university instructors based on student evaluations of their teaching effectiveness: A multilevel test of its effectiveness under consideration of bias and unfairness variables. *Research in Higher Education, 52*, 717-737. doi.org/10.1007/s11162-011-9214-7

Edström, K. (2008). Doing course evaluation as if learning matters most. *Higher education research & development, 27*(2), 95-106. doi.org/10.1080/07294360701805234

Fantini, A. E. (2018). Teacher assessment and evaluation. In The TESOL Encyclopedia of English Language Teaching (pp. 1-11). New York: John Wiley & Sons. doi.org/10.1002/9781118784235.eelt0652

Farrell, T. S. (2015). It's not who you are! It's how you teach! Critical competencies associated with effective teaching. *RELC Journal, 46*(1), 79-88. doi.org/10.1177/0033688214568096

Fordham, S. K. (2015). *Teacher and peer written feedback in the ESL composition classroom: Appropriation, stance, and authorship*. The University of Arizona.

Frey, B. B., Schmitt, V. L., & Allen, J. P. (2012). Defining authentic classroom assessment. *Practical assessment, research & evaluation, 17*(2).

Gershenson, S., & Holt, S. B. (2015). Gender gaps in high school students' homework time. *Educational Researcher, 44*(8), 432-441. doi.org/10.3102/0013189X15616123

Golding, C., & Adam, L. (2016). Evaluate to improve: useful approaches to student evaluation. *Assessment & Evaluation in Higher Education, 41*, 1-14. doi.org/10.1080/02602938.2014.976810

Hamilton, D. F., Lane, J. V., Gaston, P., Patton, J. T., MacDonald, D. J., Simpson, A. H. R. W., & Howie, C. R. (2014). Assessing treatment outcomes using a single question. *Bone Joint J, 96*(5), 622-628. doi.org/10.1302/0301-620X.96B5.32434

Hartshorn, K. J. & Evans, N. W. (2015). The Effects of Dynamic Written Corrective Feedback: A 30-Week Study. *The Journal of Response to Writing, 2*, 6-34.

Hartshorn, K. J., Evans, N. W., & Tuioti, E. A. (2014). Influences on teachers' corrective feedback choices in second language writing. *Writing and Pedagogy, 6*, 251-282.

Hendry, G. D., Lyon, P. M., & Henderson–Smart, C. (2007). Teachers' approaches to teaching and responses to student evaluation in a problem–based medical program. *Assessment & Evaluation in Higher Education, 32*, 143-157. doi.org/10.1080/02602930600801894

Hill, H., & Grossman, P. (2013). Learning from teacher observations: Challenges and opportunities posed by new teacher evaluation systems. *Harvard Educational Review, 83*(2), 371-384. doi.org/10.17763/haer.83.2.d11511403715u376

Howard, A., & Donaghue, H. (2105). *Teacher evaluation in second language education*. New York: Bloomsbury.

Hunt, K., Gurvitch, R., & Lund, J. L. (2016). Teacher Evaluation: Done to You or with You? *Journal of Physical Education, Recreation & Dance, 87*, 21-27. doi.org/10.1080/07303084.2016.1226215

Kember, D., Leung, D., & Kwan, K. (2002). Does the use of student feedback questionnaires improve the overall quality of teaching? *Assessment and Evaluation in Higher Education, 27*, 411-425. doi.org/10.1080/0260293022000009294

Ketabi, S., & Ketabi, S. (2014). Classroom and formative assessment in second/foreign language teaching and learning. *Theory and Practice in Language Studies, 4*, 435. doi.org/10.4304/tpls.4.2.435-440

Krol, M. W., Boer, D., Delnoij, D. M., & Rademakers, J. J. (2015). The Net Promoter Score-an asset to patient experience surveys? *Health Expectations, 18*(6), 3099-3109. doi.org/10.1111/hex.12297

Lang, J. W. B., & Kersting, M. (2007). Regular feedback from student ratings of instruction: Do college teachers improve their ratings in the long run? *Instructional Science, 35*, 187-205. doi.org/10.1007/s11251-006-9006-1

Leisen Pollack, B., & Alexandrov, A. (2013). Nomological validity of the Net Promoter Index question. *Journal of Services Marketing, 27*(2), 118-129. doi.org/10.1108/08876041311309243

Leung, C. (2012). Outcomes-based language teaching. In A. Burns & J. C. Richards (Eds.), *The Cambridge Guide to Pedagogy and Practice in Language Teaching* (161-179). New York: Cambridge University.

Liyanage, I., & Bartlett, B. J. (2010). From autopsy to biopsy: A metacognitive view of lesson planning and teacher trainees in ELT. *Teaching and Teacher Education, 26*, 1362-1371. doi.org/10.1016/j.tate.2010.03.006

Maltese, A. V., Tai, R. H., & Fan, X. (2012). When is homework worth the time? Evaluating the association between homework and achievement in high school science and math. *The High School Journal, 96*, 52-72. doi.org/10.1353/hsj.2012.0015

Mandal, P. C. (2014). Net promoter score: a conceptual analysis. *International Journal of Management Concepts and Philosophy, 8*(4), 209-219. doi.org/10.1504/IJMCP.2014.066899

Mark, J. & Goldberg, M. A. (2001). Multiple regression analysis and mass assessment: A review of the issues. *The Appraisal Journal*, 89-109.

Martin, R. L. (2011). The innovation catalysts. *Harvard Business Review, 89*(6), 82-87.

Mathis, W. (2012). Research-Based Options for Education Policy Making: Teacher Evaluation. National Education Policy Center. http://nepc.colorado.edu/files/pb-options-1-teval_0.pdf

McGowan, W. R., & Graham, C. R. (2009). Factors contributing to improved teaching performance. *Innovative Higher Education, 34*, 161-171. doi.org/10.1007/s10755-009-9103-6

McMillan, J. H. (2013). *Classroom Assessment: Pearson New International Edition: Principles and Practice for Effective Standards-Based Instruction*. Pearson Higher Ed.

McReynolds, J. C. (2010). Do Try this at Home: Motivating Second-grade ESL Students to Understand the Importance of Homework Completion to Improve Their Learning (Doctoral dissertation, Gratz College).

Milkova, S. (2012). Strategies for effective lesson planning. Center for Research on learning and Teaching, 1-4.

Murray, D. E., & Christison, M. (2010). *What English language teachers need to know II: facilitating learning*. Routledge. doi.org/10.4324/9780203846292

Nasser, F., & Fresko, B. (2001). Interpreting student ratings: Consultation, instructional modification, and attitudes towards course evaluation. *Studies in Educational Evaluation, 27*, 291-305.doi.org/10.1016/S0191-491X(01)00031-1

National Commission on Excellence in Education, (1983). A nation at risk: The imperative for educational reform. Washington, DC: US Government Printing Office

No Child Left Behind Act, 20 § U.S. C. 6301 (2002).

Orlich, D. C., Harder, R. J., Trevisan, M. S., Brown, A. H., & Miller, D. E. (2016). *Teaching strategies: A guide to effective instruction*. Cengage Learning.

Oxford, R.L. (1999). Anxiety and the language learner: new insights. In J. Arnold (Ed.), *Affect in Language Learning* (pp. 58-67). Cambridge: Cambridge University Press.

Pang, M. (2016). Pedagogical reasoning in EFL/ESL teaching: revisiting the importance of teaching lesson planning in second language teacher education. *TESOL Quarterly, 50*, 246-263. doi.org/10.1002/tesq.283

Paulsen, M. B. (2002). Evaluating teaching performance. New Directions for Institutional Research, 2002(114), 5-18. doi.org/10.1002/ir.42

Pennington, M. C., & Young, A. L. (1989). Approaches to faculty evaluation for ESL. *TESOL Quarterly, 23*(4), 619-646. doi.org/10.2307/3587535

Penny, A. R., & Coe, R. (2004). Effectiveness of consultation on student ratings feedback: A meta-analysis. *Review of Educational Research, 74*, 215-253. doi.org/10.3102/00346543074002215

Pratt, D. D. (2002). Good teaching: One size fits all? New directions for adult and continuing education, 2002(93), 5-16. doi.org/10.1002/ace.45

Reichheld, F. (2006). *The Ultimate Question: driving good profits and true growth.* Harvard Business Press: Boston, MA.

Reichheld, F. F. (2003). The one number you need to grow. *Harvard business review, 81*(12), 46-55.

Reichheld, F. F., & Markey, R. (2011). *The Ultimate Question 2.0: How net promoter companies thrive in a customer-driven world*. Harvard Business Press.

Richards, J. C. (2013). Curriculum approaches in language teaching: Forward, central, and backward design. *Relc Journal, 44*(1), 5-33. doi.org/10.1177/0033688212473293

Rossiter, M. J., Derwing, T. M., Manimtim, L. G., & Thomson, R. I. (2010). Oral fluency: The neglected component in the communicative language classroom. *Canadian Modern Language Review, 66*(4), 583-606. doi.org/10.3138/cmlr.66.4.583

Rucinski, M., & Diersling, C. (2014). America's teacher evaluation system revolution. Harvard Political Review. http://harvardpolitics.com/united-states/americas-teacher-evaluation-system-revolution/

Sawchuk, S. (2015). Teacher evaluation: An issue overview. Education Week, 35(3).

Sawyer, M. R., Andzik, N. R., Kranak, M. P., Willke, C. P., Curiel, E. S., Hensley, L. E., & Neef, N. A. (2017). Improving Pre-Service Teachers' Performance Skills Through Behavioral Skills Training. *Behavior Analysis in Practice*, 1-5. doi.org/10.1007/s40617-017-0198-4

Schneider, D., Berent, M., Thomas, R., & Krosnick, J. (2008, June). Measuring customer satisfaction and loyalty: Improving the 'Net-Promoter' score. Presentation at the Annual Meeting of the American Association for Public Opinion Research, New Orleans, Louisiana.

Shin, T. W., & Manochphinyo, A. (2017). The Perceptions of Malaysian Students toward Native and Non-native ESL Teachers. *The English Teacher, 46*(1), 17.

Simpson, P., & Siguaw, J. (2000). Student evaluations of teaching: An exploratory study of the faculty response. Journal of Marketing Education, 22, 199-213. doi.org/10.1177/0273475300223004

Spencer, K. J., & Schmelkin, L. P. (2002). Student perspectives on teaching and its evaluation. Assessment & Evaluation in Higher Education, 27(5), 397-409. doi.org/10.1080/0260293022000009285

Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching: The state of the art. *Review of Educational Research, 83*, 598-642. doi.org/10.3102/0034654313496870

Su, T., & Tian, J. (2016). Research on corrective feedback in ESL/EFL classrooms. *Theory and Practice in Language Studies, 6*(2), 439. doi.org/10.17507/tpls.0602.29

Tan, F., Nabb, L., Aagard, S., & Kim, K. (2010). International ESL graduate student perceptions of online learning in the context of second language acquisition and culturally responsive facilitation. *Adult Learning, 21*(1-2), 9-14. doi.org/10.1177/104515951002100102

Thomsen, J. (2014). A Closer Look: Teacher Evaluations and Reduction-in-Force Policies. Education Commission of the States.

Tsiplakides, I. and Keramida, A. (2010). Promoting positive attitudes in ESL/EFL classes. *The Internet TESL Journal, XVI*(1).

Tsui, A.B.M. (1996). Reticence and anxiety in second language learning. In K.M. Bailey and D. Nunan (Eds.), Voices from the Language Classroom: Qualitative Research in Second Language Education (pp. 145-167). New York: Cambridge University Press.

Vu, M. T. (2016). The kaleidoscope of English language teacher professionalism: A review analysis of traits, values, and political dimensions. *Critical Inquiry in Language Studies, 13*(2), 132-156. doi.org/10.1080/15427587.2016.1146890

Walberg, H.J., Paschal, R.A., & Weinstein, T. 1985, 'Homework's Powerful Effects on Learning', *Educational Leadership, 42*(7), pp. 76-79.

Walkinshaw, I., & Oanh, D. H. (2014). Native and non-native English language teachers: Student perceptions in Vietnam and Japan. *SAGE Open, 4*(2), 2158244014534451. doi.org/10.1177/2158244014534451

Wallinger, L.M. 2008, 'The Role of Homework in Foreign Language Learning', *Foreign Language Annals, 33*(5), pp. 483-496. Abstract doi.org/10.1111/j.1944-9720.2000.tb01993.x

Walsh, S. (2006). Talking the talk of the TESOL classroom. *ELT Journal, 60*(2), 133-141. doi.org/10.1093/elt/cci100

Wilkinson, L., & Dallal, G.E. (1981). Tests of significance in forward selection regression with an F-to enter stopping rule. *Technometrics, 23*, 377-380. doi.org/10.1080/00401706.1981.10487682

Young, D.J. (1991). Creating a low-anxiety classroom environment: What does language anxiety research suggest? *The Modern Language Journal, 75*(iv), 426-439. doi.org/10.2307/329492

About the Authors

K. James Hartshorn has been involved in second language education in the United States and Asia for three decades. James currently serves as a faculty member in the Linguistics Department at Brigham Young University where he mentors new TESOL professionals by helping them optimize the efficacy of their ESL teaching and research.

Judson M. Hart holds an MA in TESOL and currently serves as Director of Language Assessment at Emmersion.

Benjamin L. McMurry has nearly 20 years of experience working with non-native English speakers. He currently serves as the Program Coordinator at the English Language Center at Brigham Young University. His research interests include materials development and evaluation.

Appendix A

Principled Pedagogical Practices

1. Rely on course outcomes

Teachers understand the course outcomes for the skill and proficiency level in which they teach and effectively communicate them to students. They can describe student behaviors that demonstrate these outcomes, and they successfully design classroom-learning activities that help students progress toward achieving them. Teachers engage in ongoing informal and formal assessment activities and provide personalized feedback based on the course outcomes.

2. Plan lessons effectively

Teachers carefully plan lessons so language development will be optimized during the class period. Teachers plan to incorporate an appropriate number and variety of learning activities that are meaningful and engaging. These activities build incrementally from more simple uses of language to more complex uses that are authentic and communicative. Teachers consider the best ways to ensure that communication of explanations and expectations are clear and concise in order to maximize student language practice. This includes preparing the board or other materials well ahead of class time. Teachers also prepare contingency plans in order to adjust for a variety of unforeseen circumstances and changing student needs.

3. Optimize class time

Teachers feel a sense of urgency about using as much of the classroom time as possible for meaningful language practice. They convey this sense of urgency to their students by starting class on time and by carefully managing activities and transitions in order to maximize communicative language practice. However, rather than rushing through their lessons, teachers skillfully connect activities and ensure that students achieve the needed level of mastery before moving on. They anticipate potential threats to effective use of class time such as problems with technology, excessive student questions, inappropriate student behaviors and so on. Their responses to such challenges are principled and appropriately bring the class back on course. Teachers also end class on time.

4. Cultivate a positive learning environment

Teachers understand the necessity of a positive learning environment in order to optimize learning. They recognize that positive teacher-student interaction is at the heart of the environment they seek to cultivate. They foster genuine concern for their students and their learning based on principles of respect and trust. They leave personal concerns behind as they plan and teach their classes. They are consistent and equitable in their classroom practices and help students to see how classroom policies and activities facilitate language development. They create a non-threatening learning environment that is cheerful, upbeat, and optimistic. They inspire students to do their best, and they help them experience the joy of effectively applying what they learn. They sincerely praise students and regularly express confidence in their abilities.

5. Evaluate learning effectively

Teachers are committed to the ongoing evaluation of student learning. They skillfully use diagnostic tests, classroom instruction, language practice, and formal and informal assessments to clarify individual learner needs in relation to established course outcomes. They also regularly solicit qualitative input from their students regarding learning materials and methods. This information is then used to make appropriate adjustments in lesson planning and the selection of materials and methods used in the classroom. Teachers help students to understand the rationale for adjustments that are made as well as areas where continuity may be necessary.

6. Utilize homework strategically

Teachers understand the potential for effective homework to help students achieve course outcomes. Rather than assigning busy work, they carefully consider the quantity and specific kinds of learning activities that are needed by their students in order to foster language development or to help them better understand and diagnose learner needs. They are able to effectively communicate the rationale for various types of homework to their students. They demonstrate the value of the homework in the way they follow up and process the homework. They know when it may be appropriate to review certain types of homework in class and when the class time should be used for other activities. They utilize student performance on homework to inform their ongoing instruction in the classroom.

7. Provide meaningful and timely feedback

Teachers know that feedback is essential to effective learning. They regularly provide students with feedback that is meaningful—it focuses on the most important language elements for each learner; students understand the feedback, why it was given, and how to apply it. Though teachers ensure that ongoing feedback is timely, they are careful not to overload the students' cognitive ability to process and apply the feedback. Along with feedback, teachers provide students with abundant opportunities to practice and apply the feedback in a variety of learning contexts.

8. Exemplify professionalism

Teachers value and participate in orientations, training, and workshops. They are well prepared, punctual, and complete all administrative tasks on time. They act and look the part of a professional in the classroom including adhering to the dress and grooming standards and maintaining appropriate teacher-student boundaries. They are respectful and courteous with their students and other teachers with whom they share resources such as classrooms, offices, technologies, and learning materials. They consistently evaluate their own teaching and seek to improve through feedback from students, administrators, and peers. They appropriately apply the relevant feedback they receive.

Appendix B

Survey Components Presented to Students

My [Course] teacher, [Teacher]

	Rarely	Sometines	Usually	Almost Always	Always
teaches toward course outcomes.	○	○	○	○	○
plans lessons effectively.	○	○	○	○	○
gets the most from each class hour.	○	○	○	○	○
creates and keeps a positive learning environment.	○	○	○	○	○
evaluates learning effectively.	○	○	○	○	○
gives useful homework.	○	○	○	○	○
provides helpful feedback quickly and frequently.	○	○	○	○	○
is a good example of a professional teacher.	○	○	○	○	○

Group Listening Quizzes

Veronica Wright, Brigham Young University–Hawaii, Laie, Hawaii, USA

Introduction

The power of collaborative work among learners has been well documented in pedagogical research. Starting with Vygotsky's (1962) theory of Zone of Proximal Development, which posits the idea that learners are able to develop a greater range of skill with the help of peer collaboration or adult guidance than they can on their own, cooperative learning methods continued to flourish across a wide range of educational fields. At the turn of the century, Johnson, Johnson, and Stanne (2000) completed a meta-analysis of over 900 research studies that reported on cooperative learning methods, concluding that all of the studies they reviewed validate the effectiveness of cooperative learning over individualistic or competitive learning. Studies such as the one performed by Ravenscroft et al. (1995) corroborate these findings by showing that learners who were graded on team effort as well as individual effort performed better on exams than those in a control group that were graded solely on individual effort. Additionally, Clinton and Kohlmeyer (2005) reported a variety of positive effects of group work that manifested in learners some of which included significantly increased motivation to learn, increased enthusiasm, a belief that they had improved their problem solving abilities, higher ratings of their instructor's overall performance, and a belief that the instructor had helped them to learn more than those who were in the group that did not use group quizzes as part of their instruction. Due to its numerous effective benefits, group work has permeated ESL and EFL instruction for many decades. It has been included in the instruction of almost every skill. One such skill area that can greatly benefit from collaborative work is listening.

English language learners (ELL) at the university level have demanding expectations placed on them when it comes to the skill of listening. Once these learners enter university level classes, they are expected to listen to lectures, watch videos, and interact with native English speakers. Flowerdew and Miller (2014) enumerate the listening difficulties ELL learners typically face in university such as speed of lectures, speakers' accents, vocabulary load, identifying the lecture structure and organization, and many others. Much of the information learners listen to is essential to their learning of the subject and therefore helping students develop their listening comprehension skills in English for Academic Purposes (EAP) listening and speaking classes is vital. Assessing listening comprehension to track and improve learners' progress in an EAP classroom can be difficult for a variety of reasons. One such reason is that we actually know relatively little about the processes students use when they attend a lecture in a second language (Field, 2011). Another such difficulty in teaching and assessing listening comprehension occurs because "the processes that learners engage in during listening cannot be directly observed and controlled" (Goh, 2014, p. 72). For example, when using quizzes with multiple choice questions to assess learners' comprehension, instructors may be unable to understand why learners chose the answers they did out of a list of possible choices. One potential solution for this is to have learners explain why they chose the answers they did but this is not part of the traditional practice of using multiple choice quizzes. This teaching tip discusses how group quizzes can be leveraged to assess listening comprehension in a way that allows students to access the many benefits of collaborative learning and provides us a glimpse into what is happening inside the minds of learners.

Group Listening Quizzes

Using group listening quizzes is a way to assess listening comprehension that combines the method of using multiple choice quizzes to assess listening with the opportunity for learners to discuss answers within a group setting. In this approach, language learners in an EAP listening class are tasked with the assignment of listening to a lecture twice and taking notes. They are then given a multiple choice quiz to take which they first take individually and then again as a group. In preparation for group listening quizzes, it would be beneficial for learners to be taught and given practice in taking notes and have prior experience working with others in a group setting. Quizzes used can either be created by the teacher or taken from classroom materials or textbooks.

Materials

The materials necessary for this activity are: note-taking materials, video/audio equipment for viewing and listening to the lecture, and copies of the quiz. Copies of the quiz should include enough copies for each student to take the quiz individually and additional copies for the group quiz, one quiz sheet per group.

Procedure

Step 1: Prepare students for the lecture by explaining that they will listen to the lecture two times and that they should take notes as they listen both times. Tell students the title of the lecture and how long it is, then play the lecture twice.

Step 2: After students have listened to the lecture, pass out copies of the quiz to each student. Have students use their notes to take the lecture quiz on their own. Once they are finished, have them turn in their quiz to be graded. This individual quiz can be timed or not.

Step 3: Once students have turned in their individual quiz, put them into groups of three or four and give each group one new copy of the quiz. Allow students a limited amount of time to take the quiz again. Explain that the (entire) group will need to come to an agreement for each of the answers they choose. Encourage learners to discuss their answers and use their notes as evidence to support the answer they think is correct.

Step 4: Make sure to circulate and listen to students as they attempt to explain why they think an answer is the correct one. Listen for any break downs in listening comprehension or note taking and use that information to inform future lessons.

Step 5: Once each member has agreed on one answer for each question, have the students submit their group quiz to be graded. Make sure that the students know that they will get a grade for both quizzes (their individual and their group quiz). Teachers may opt to weight one quiz slightly more than the other or put the different quiz scores into different categories depending on the grading categories of the course. For example, I prefer to put the group quiz under the grading category of engagement which measures students' participation and the individual quiz under the listening proficiency category which measures the students' actual listening ability. This way, they will receive credit for the work they did on both quizzes and students are less likely to get upset when teammates choose incorrect answers.

Step 6: In subsequent group listening quizzes, make sure to rotate group members to provide variety and practice in negotiating meaning with different groups of peers.

Adaptations

Adaptations for group listening quizzes can be made according to class sizes, proficiency levels, and EFL or ESL contexts. For example, group sizes can be larger or smaller based on the number of students in the class. In addition, depending on students' proficiency, the teacher can increase or lessen the number of times the students are allowed to listen to the lecture and the amount of time they are given to take notes. Finally, the number of lectures and listening quizzes could increase or decrease depending on whether the class is being taught in an EFL or ESL environment to help provide enough comprehensible input.

There are also multiple adaptations that may help to increase learners' motivation. One example is to make the group quiz score worth slightly more than the score for the individual quiz. This can increase learners' motivation to do well on the group quizzes due to the higher impact the quiz will have on their grades. Another way to adapt this assignment to help increase group participation is to mix the groupings. Having students with mixed proficiency levels in each group will help scaffold the listening activity for lower proficiency learners and provide more opportunities to negotiate meaning as well as opportunities for peer mentoring. Allowing higher proficiency learners the opportunity to collaborate with lower proficiency learners can be motivating for some students (Oxford, 1997). A third way to increase motivation is to make the group quiz a game where each group competes with other groups to get 100%. To increase the stakes, a timing factor can be included. The time factor could be determined by the teacher or can be determined by which group finishes first, meaning, the time to take the quiz is done for everyone as soon as the first group finishes. To increase the spirit of competition, the answers can be reviewed as a class. Because the groups share their answers out loud in class, each group can compare their scores with other groups to see which group has achieved the 100% goal. This can possibly create a positive feeling of collaboration and accomplishment through friendly competition.

Benefits

There are multiple potential benefits of using group listening quizzes. One substantial benefit is the opportunity for students to verbalize their thinking processes. As they discuss their own thinking processes in relation to the lecture they listened to, they can compare their thinking with their peers' mental processes and fill in the gaps between the two. As teachers circulate and listen to the group discussions, they are also privy to these insights and can use them to inform and improve their teaching. Another benefit is increased participation particularly among students who have a high level of listening comprehension but are quiet and reserved. These students often will get full points on their individual quizzes but may allow louder more confident students to persuade the group to choose the wrong answers during the group quiz. Providing the opportunity to take both quizzes allows these students to see their ability in the individual quiz which could then give them the confidence to speak up during their group discussion in order to get full points on their group quiz as well.

There can also be a significant increase in negotiation of meaning among learners. The opportunity to discuss the answers in small groups allows students the opportunity to define the meanings of terms in the lecture and in the listening quiz itself as well as clarify information in their notes. This negotiation also gives learners practice in using the terminology specific to the topic in the lecture. Because the group quizzes are graded, learners have an added motivation to argue for the answers that they feel are right. Increased use of speaking strategies such as using speech acts of persuasion and argumentation can also manifest themselves in this approach because students need to learn how to argue for the answers they believe are right. Providing instruction and feedback on useful phrases for dealing with persuasion and disagreement can further student learning in this area (Wong and Waring, 2020).

After the first group quiz, learners will often see the advantages of taking good notes. Clear and thorough notes help them to answer the quiz questions and persuade their group members of the correct answers by providing evidence with which they can convince their peers. This can increase learner motivation to work on improving their notes and apply note-taking strategies learned in class.

Conclusion

Group listening quizzes can be one way to access the power of collaborative learning, thereby potentially increasing learner motivation and negotiation of meaning among learners. It can be a beneficial tool that encourages peer interaction, scaffolding of learning, and even mentorship. Learners have opportunities to develop pragmatic competence in the act of persuasion and participate in deeper cognitive processing of materials that they listened to. And finally, both teachers and students have the opportunity to get a glimpse of the mental processes that learners are using as they are listening to materials within the classroom through the conversations they have during group quizzes.

References

Clinton, B. D., & Kohlmeyer III, J. M. (2005). The effects of group quizzes on performance and motivation to learn: Two experiments in cooperative learning. *Journal of Accounting Education, 23*(2), 96-116.

Field, J. (2011). Into the mind of the academic listener. *Journal of English for Academic Purposes, 10*(2), 102-112.

Flowerdew, J., & Miller, L. (2005). *Second language listening: Theory and practice*. New York: Cambridge University Press.

Goh, C. C. M. (2014). Second language listening comprehension: Process and pedagogy. In M. Celce-Murcia, D. M. Brinton, & M. A. Snow (Eds.), *Teaching English as a Second or Foreign Language* (4th ed., pp. 72-89). Boston, MA: National Geographic Learning.

Johnson, D. W., Johnson, R. T., & Stanne, M. B. (2000). Cooperative learning methods: A meta-analysis. Retrieved September 28, 2022, from https://

www.researchgate.net/publication/220040324_Cooperative_learning_methods_A_meta-analysis

Oxford, R. L. (1997). Cooperative learning, collaborative learning, and interaction: Three communicative strands in the language classroom. *The Modern Language Journal, 81*(4), 443-456.

Ravenscroft, S., Buckless, F., McCombs, G., & Zuckerman, G. (1995). Incentives in student team learning: An experiment in cooperative group learning. *Issues in Accounting Education, 10*(1), 97–109.

Vygotsky, L. (1962). *Thought and language* (E. Hanfmann & G. Vakar, Trans.). Cambridge, MA: MIT Press. https://doi.org/10.1037/11193-000

Wong, J., & Waring, H. Z. (2020). *Conversation analysis and second language pedagogy: A guide for ESL/EFL teachers*. New York: Routledge.

About the Author

Veronica Wright teaches English as an international language and TESOL courses in the English Language Teaching and Learning Program as an assistant professor at Brigham Young University–Hawaii. Her other interests include vocabulary learning, learner corpora, and the use of corpora to teach grammar principles.

Teaching and Assessing Writing Using a Triage Approach

Leola Solis, Brigham Young University–Hawaii, Laie, Hawaii, USA

Introduction

The word *triage* is a term used in the medical field to explain the "sorting of and allocation of treatment to patients…according to a system of priorities designed to maximize the number of survivors" (Merriam-Webster, 2022). The use of this word in this teaching tip denotes a prescriptive approach to assessing writing. The teacher is the doctor, the students are the patients, and the writing errors are the symptoms that need to be addressed by being sorted, prioritized, and treated. The Triage approach to a student's writing process is a way to help students notice (Schmidt, 1990), recognize, and apply an appropriate treatment for their writing weaknesses.

This teaching tip can be conducted within a variety of learning situations. It can be applied to English as a Second Language (ESL) or English as a Foreign Language (EFL) learning situations that fit the post-pandemic era we currently live in or at the other end of the spectrum, a well-funded English for Academic Purposes (EAP) institution that includes tutors and language centers for learning. This article will focus primarily on explaining how to implement this teaching tip in a post-pandemic ESL environment where resources are limited.

The teaching tip is simple in its delivery. The teacher assigns students a topic to write about. Each student writes on the assigned topic, reviews their own work and a peer's work, identifies writing errors within their own work and a peer's work, compiles a list of writing errors which also includes errors listed from the teacher's review, prioritizes what errors need treatment first, and then completes the treatment process using a three-pronged approach. The complete teaching tip can be repeated as many times as the teacher sees the need. Using this teaching tip will help students become better writers and build within them an awareness of their own writing strengths and weaknesses.

Literature Review

Much has already been written about corrective feedback, self-assessment and peer-assessment in L2 writing. There is evidence to suggest that in regards to corrective feedback, scaffolding the feedback for individual learners helps learners become more self-regulated and independent learners (Bitchener, 2012). Andrade et al. (2007) found that self-assessment or self-review not only helped the learners align expectations with the teacher, especially with the help of a rubric, but also eased the learning process. Other studies that focused on combining the self-correction and peer-correction techniques found that using both techniques simultaneously had a significant impact on reducing student errors and improving the quality of their writing (Yanti, et al., 2022; Ganji, 2009).

In addition to looking at research studies of the benefits of the different feedback techniques, it is important to define the principles of "noticing" and "self-regulated learning" since they are principles that this teaching tip shows as possible products of its implementation. Schmidt's (1990) "noticing hypothesis" posits that learning takes place when learners become aware of their errors. In light of this, this teaching tip is designed to help learners become more aware of their writing errors during the different assessment processes. As for "self-regulated learning," Pintrich (1995) suggests that a self-regulated learner is one who acts for themselves, has a goal in mind, and controls their own behavior, motivation, emotions, and thinking. Following the steps of this teaching tip, which seeks to combine the techniques of corrective feedback, self-assessment and peer-assessment for assessing writing, the principle of "noticing" as it pertains to language acquisition (Schmidt, 1990), and the principle of self-regulated learning as defined by Pintrich (1995), will assuredly create self-regulated and independent learners with the tools necessary to becoming more proficient and skilled writers.

Procedure:

Step One: In-class writing assignment: Present the class with the first in-class writing topic. The topic should be something that students have background knowledge in to trigger schemata; it can also be on a topic that is currently being discussed in class. The topic should fit the proficiency of the students. If students have lower proficiency in English, the teacher can assign a simple topic such as "*why do you want to learn English?*" If the student is at a higher proficiency, the topic can be a little more complex; for example, "*write about three important lessons that you have learned in your lifetime about relationships or about yourself.*" For a higher academic English course the topic can be related to what is being discussed in class; for example, "*In our discussion of the "Allegory of the Cave" we talked about 'modern caves' in our life. In a four to five paragraph essay explain how culture can be a "cave." Explain what you think some chains are that bind you to this cave.*" The length of the writing sample could be three to five sentences or a paragraph for lower proficiency students or a three to five paragraph essay for higher proficiency students. The writing can be timed to check for automaticity of the language or untimed depending on the needs of the students.

Step Two: Three Reviews and Identifying Writing Errors: After the writing activity is completed, the next phase of the triage approach begins. This phase deals with identifying what writing errors students are making in order to prescribe treatment. This consists of completing three reviews: self-review, peer-review (or online review), and teacher review. For each review, a writing review checklist (Appendix A) is completed.

For the self-review assignment, students first complete the writing review checklist (Appendix A). This checklist draws their attention to strengths and weaknesses in their writing style. They then complete a second checklist, the grammar review checklist (Appendix B) by marking in the right-hand column the number of times each grammar error occurs. This helps them to identify common grammar errors within their writing sample. For the peer-review process, assign students a partner. Each partner will read the other partner's writing sample, complete the writing review checklist (Appendix A), and then discuss with the writer any comments they made as they completed the review checklist. The peer-review assignment may take a little more teacher planning and effort if the class is being taught remotely. For the peer-review assignment in an online class, the teacher can assign each student a partner and connect each student with their partner through email, messenger, or any one of the online video applications available to students, such as, Zoom, google meet, skype, etc. If it is difficult for a student to meet with a peer using one of the online applications listed above because of problems with low bandwidth, different time zones, or lack of resources, they could be directed instead to online resources such as Grammarly, ProWritingAid, and or WhiteSmoke, etc., for more input about their writing. Using one of these online resources would take the place of the peer-review. Some in-class teaching activities surrounding these online resources may be needed to help students become more familiar with the use of these resources. There are limits to what technology can do in comparison to a teacher; however, online software have been found to be effective tools for students editing their own mistakes (Cowan et al., 2014). Finally, the last review in the "three reviews" step is completed by the teacher. The same pattern can be applied to the teacher review. The teacher could meet with the students to discuss their writing strengths and weaknesses according to both the writing review checklist (Appendix A) and the grammar review checklist (Appendix B)

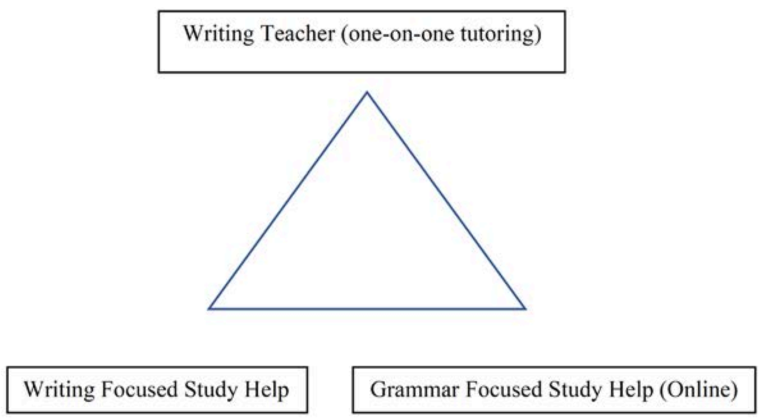
At the completion of the "three reviews" process, the student should have within their possession three writing review checklists and two grammar review checklists. From these checklists, they can complete a triage of their writing weaknesses. For the writing review checklists, students will look at all the errors identified in the three checklists and then identify the one that occurs the most. They will then focus on improving this error in step three. For the grammar review checklists, the students will look at all the grammar errors listed in the two grammar review checklists and count how many times a grammar error was committed. They then choose the grammar error that had the highest count as their focus for the treatment process.

Step Three: The three-pronged approach: After the writing and grammar errors have been identified at the end of step two, the treatment is prescribed and step three begins. The treatment is a three-pronged approach. In this three-pronged approach, students will do three things to address their immediate writing and grammar problems. These three things can be adapted according to what resources are available to students. In this post-pandemic era where resources have become more readily available online (e.g., The Purdue Online Writing Lab, Grammarly, Grammar Girl), and remote conferencing is much more streamlined (e.g., Zoom, Google Meet, Skype), the three-pronged approach could be the following: meet with writing teacher for one-on-one tutoring (instead of the tutor), study a writing strategy online (instead of attending a workshop at the institution's learning center), and work on a grammar weakness online (instead of at the institution's learning center).

Three-pronged approach for an ESL/EFL writing class with limited resources

Figure 1 shows a visual depiction of the three-pronged approach described in this teaching tip.

Figure 1. Three-pronged approach



Writing Teacher (one-on-one tutoring):

For the first part of the three-pronged approach, students can either meet with a writing tutor, if one is available, or meet with their writing teacher. This may mean more work for the teacher; however, the benefits of a one-on-one session with students is very valuable to a student's learning (Grasha, 2010). To prepare for the tutoring session, students should have with them the following: their writing sample, all of their completed checklists from step two, and an idea of what writing and grammar weaknesses to discuss with their teacher. The writing and grammar weaknesses should have been chosen from step two. It should be the most frequently noted errors written on their checklists from the self-review, peer-review (online review), and teacher review.

During the tutor session, the tutor (or teacher) will go over the student's writing sample first with the student, then look over the checklists the student brought to the appointment. The tutor will ask the student what topic they wish to discuss (giving students some control over the appointment) and then proceed from there. They will then discuss a few writing and grammar tips based on the topics of choice to help the students with their writing and grammar. The suggested timing of the session is 30 minutes with a possible extension on time if the need arises.

Writing Focused Study Help:

The next two prongs of the approach are made up of the writing and grammar focused study help. These two areas are separated for the purpose of this teaching tip to allow for a more focused form of study. The topic for the writing focused study should concentrate on the conventions of writing (i.e., structure, topic sentences, style, etc.); for example, if the writing objective was a paragraph, did the student follow the paragraph structure and include a topic sentence, supporting details, and a conclusion. If the writing objective was an essay, did the student follow the essay structure and include a thesis statement, introduction, correct rhetorical pattern, etc. In an environment with limited resources, the teacher could record a workshop on a specific topic and have students watch the recording and complete some assignments on the assigned topic at a later time outside of class. As the teacher, you could also provide online resources for students to study. One such online resource is: Purdue Online Writing Lab (OWL Purdue, 2022). This website has resources on academic writing, writing style, essay writing, paragraphs and paragraphing, etc. This is a great resource for students as it is rich in useful writing information.

Grammar Focused Study Help:

The grammar focused study help is a companion to the writing focused study help. The grammar focused study help can include topics such as subject verb agreement, verb tense, articles, modifiers, pronouns, prepositions, etc. It is also recommended that students choose the topic for this section from the feedback they received from the two reviewers in step two of the triage approach (see Appendix B). As the teacher, you could create a list of online study help resources for students to choose from based on the grammar review checklist (Appendix B). Purdue Online Writing Lab (OWL Purdue) provides study help on many grammar topics for English as a Second Language Learners. Some of the topics listed on the OWL Purdue website specific to grammar are: Combining Sentences, Nominalizations and Subject Position, Prepositions, Pronouns, Punctuation, Paraphrasing and Summary, etc. The website also includes OWL Exercises such as: Grammar Exercises, Punctuation Exercises, Spelling Exercises, Sentence Structure, Sentence Style, Writing Numbers, and ESL Exercises. This website is very useful not only for writing focused study help, but also for the grammar focused study help angle of the three-pronged approach as well. Other resources available for grammar study help can be found on YouTube such as, JenniferESL, Bob the Canadian, or EF podEnglish, etc. It is recommended that the teacher study these videos beforehand to help tailor the online resources to the student's needs. Resources available online target different proficiency levels and different topics so finding what fits your student's proficiency and grammar needs best would make this treatment of their writing errors more effective.

The purpose of the writing focused and grammar focused study help is twofold, it provides learners with online resources that they can revisit at any time during their studies thus creating self-regulated learners, and it strengthens or solidifies the information that the student has learned from their teacher during the writing teacher session. Repetitiveness of key concepts can eventually lead to automaticity because of familiarity. Having multiple points of contact on a subject can also act as "checks and balances" for the student so they are hearing about the same topic from multiple experts rather than just one.

Step Four: Reporting: After students have completed the three-pronged approach, they will write a report (in essay format) detailing what they did for each of the three points in the triangle. This helps the student to track their learning process similar to what journaling does, and gives them another chance to revisit what they have learned and hopefully display their learning through their writing. The first paragraph will state the topic of the in-class essay, the feedback they received from the reviews, and a list of the three things they did to complete the Triage. In the body of the essay, they should describe their tutor session (or teacher session), what grammar weakness they studied, and what they learned in the writing study help website or workshop. The last paragraph will be what they learned from this Triage process and whether they feel there was any improvement in their writing. From these reports, the teacher can assess the student's writing and examine how well they have applied what they have learned from the Triage assignments. The reporting step of this teaching tip is important for students. It is a self-reflection of their own writing which is integral to the learning process as it allows the student to keep track of their progress and growth in their writing.

Grading:

Initially, for the in-class writing activities, grading is pass or no pass. Students must complete every step of the Triage process to pass. Grading writing assignments using a rubric is pushed to later in the school year when the teacher feels students understand the rubric, the expectations of the teacher, the course, and the institution. Because of the Triage assignments and the feedback provided from their peers, tutors, and teacher, students are better prepared and more aware of the expectations of the teacher and of the writing course. It is at this point, as students' writing and the teacher's expectations begin to align together, that grading begins in earnest.

Preparation

To prepare for this teaching tip, you will need to find instructional videos online that provide specific writing instruction based on the writing review checklist (Appendix A). You will also need to find grammar quizzes or instructional grammar videos of common grammar errors that students make based on the grammar review checklist (Appendix B). Collecting these online instructional help videos will make it easier for students to complete the writing focused and grammar focused study section of step three. You will also need to prepare paragraph or essay topics for the timed in-class writing activities.

Materials

To complete the "three reviews" step, you will need to provide a writing and grammar checklist for learners (Appendices A and B). These checklists can be long and exhaustive, or simplified to the goals of each writing task. For the teacher review, if you have a rubric that you would like to use, you could use this to complete the teacher review (see Appendix C). You could also use Appendices A and B for your teacher review. Using the same checklists for all reviews in step two will give students a clear picture of what others see as their most frequent writing errors. They can then compare this to their own self-review checklist. It can also lead to a clearer understanding of the student's writing and grammar weaknesses and places the treatment in step three directly where it belongs, on the areas that need fixing.

Conclusion

This teaching tip works best if it is repeated throughout the length of the course. In this way students are working on different writing and grammar study helps and seeing a tutor or the teacher multiple times throughout the length of the course. Although the three parts of the three-pronged treatment (writing teacher/tutor, writing focused study help, and grammar focused study help) never change each time students complete this teaching tip, what students do specifically with each treatment does change. Each time they see the writing teacher or tutor, they are discussing something specific to them that was decided on after reviewing the self-assessment and the peer-review. Each time they choose a specific grammar rule to study, they are focusing on a different area of weakness in grammar. The three prongs of the approach have purposely been chosen, not only to help students improve their writing skills, but also to help them become self-regulated learners.

References

Andrade, H., & Du, Y. (2007). Student responses to criteria–referenced self–assessment. *Assessment & evaluation in higher education, 32*(2), 159-181.

Bitchener, J. (2012). Written corrective feedback for L2 development: Current knowledge and future research. *Tesol Quarterly, 46*(4), 855-860.

Cowan, R., Choo, J., & Lee, G. S. (2014). ICALL for improving Korean L2 writers' ability to edit grammatical errors. *Language Learning & Technology, 18*(3), 193-207.

Ganji, M. (2009). Teacher-correction, peer-correction and self-correction: Their impacts on Iranian students' IELTS essay writing performance. *Journal of Asia TEFL, 6*(1).

Grasha, A. F. (2002). The dynamics of one-on-one teaching. College Teaching, 50(4), 139-146.

Merriam-Webster. (n.d.). Triage. In *Merriam-Webster.com dictionary*. Retrieved July 25, 2022, from https://www.merriam-webster.com/dictionary/triage.

Pintrich, P. R. (1995). Understanding self–regulated learning. *New directions for teaching and learning, 1995*(63), 3-12.

Purdue Writing Lab. (n.d.). Purdue OWL // Purdue Writing Lab. Retrieved July 25, 2022, from https://owl.purdue.edu/owl/purdue_owl.html

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics,* 11, 129–58.

Yanti, A., Hadi, A., & Settiawan, D. (2022). Using Self-Correction and Peer-Correction Techniques to Improve Students' Writing Skills: A Comparative Study. *Indonesian Journal of Integrated English Language Teaching, 8*(1), 1-14.

About the Author

Leola Solis has been teaching English as a second language for 20 years. She is an Assistant Professor at Brigham Young University - Hawai'i and a graduate of the MATESOL program at Hawai'i Pacific University.

Appendix A

Sample Writing Review Checklist

Peer Review Checklist Handout Name: ______________

Peer Review One: Introductions

______ Does the introduction effectively introduce the topic or problem, not just jumping to the thesis?
______ Does it provide necessary background (history, definition, description, factual/historical, shared experience, narration, theory)?
______ After providing effective background and creating a need for the thesis, does the writer effectively introduce and state that thesis?
______ Is the thesis stated in one sentence?
______ Does the thesis address the topic directly and limit the topic with a central idea (an opinion, idea, or attitude about the topic)?
______ Does the thesis present or hint at the logical development of the paper? If not, is it followed by an explicit statement of development to guide the reader through the paper?

Peer Review Two: Essay Structure

______ Does the paper follow the organization hinted at in the introduction?
______ Do the topic sentences provide the main idea for each paragraph, have a controlling idea, and clearly relate to the thesis statement?
______ Do paragraphs flow naturally from each other? Do the paragraphs seem to appear in a logical, understandable order? Will they smoothly and convincingly move the reader from the introduction to conclusion?
______ Does the essay effectively use an appropriate rhetorical pattern?
______ Does the conclusion go beyond a restatement/summary to bring the essay to a satisfying end using an effective style (i.e., recommendation, prediction, quotation, result), and avoid new support or ideas?
______ Is the essay unified with information in each paragraph directly related to the topic sentence and thesis?

Peer Review Three: Framing Method

In the body paragraphs, does the writer 'frame' the paragraph in this way:
______ begins with a topic sentence?
______ introduces the source?
______ Uses the source? (whether paraphrase, quote, statistic, whatever)
______ Cites the source? (don't worry about this step right now)
______ Comments on it before moving on?
______ Repeats the above steps if another source is introduced?
______ Closes the paragraph with a reiteration of the topic sentence-though not in the same words?

Peer Review Four: Style and Conciseness

______ Does the essay use appropriate style?
______ Does the essay avoid announcing or the use of unnecessary words or phrases such as: *I think, I believe, In my opinion…?*
______ Does the essay communicate ideas clearly with consistent language?
______ Does the writer use vocabulary that shows variety and consistent precision in form and meaning?
______ Do the sentence structures show variety, complexity, and sophistication?

Appendix A: Used with permission

Appendix B

Sample Grammar Review Checklist

Name: ________________________ Essay # ______________

Symbol Chart for Grading Comments		
Symbol	**Writing Concern**	**A Problem in Your Paper**
¶	New paragraph	
cap	Capitalization	
ss	Sentence Structure	
inc	Fragment (Incomplete Sentence)	
ro	Run-on sentence	
…	Add more details	
?	Meaning not clear	
redun	Redundant	
wdy	Wordy (unnecessary words)	
∧	Missing a word	
	Omit word	
( )	Words not needed/wrong	
	Word order	
wf	Word form	
vf	Verb form	
wc	Word choice	
vt	Verb tense	
pp	Preposition (missing/wrong)	
art	Article (missing/wrong)	
pn	Pronoun (missing/wrong)	
sp	Spelling	
s/v	Subject – Verb Agreement	
s/pl	Singular – Plural Agreement	
c/nc	Count/non-count error	
"…"	Quotation Marks	
App	Apostrophe	
c	Comma	
cs	Comma splice	
;	Semicolon	

Appendix B: Used with permission

Appendix C

Sample Writing Rubric

Essay Rubric (Adapted from the CSU English Placement Test (EPT Scoring Guide, was created by Mary Wardell and Silverio Haro; http://www.csusm.edu/lwap/lessessayrubric.html)

	A. Command of Topic	
5	Essay addresses the assignment, demonstrating both understanding of the literal meaning of the text and the ability to interpret meaning beyond the literal meaning.	
4	Essay addresses the assignment, showing understanding of the literal meaning of the text.	
3	Essay attempts to address the assignment, but misses some element of the literal meaning of the text.	
2	Subject is generally clear but does not always keep to assigned topic.	
1	Essay does not directly address assigned topic.	
	B. Argument Development	
5	Essay convincingly asserts an analytical thesis and carefully develops related ideas in coherent, sequential paragraphs, providing convincing textual and other evidence.	
4	Essay asserts a thesis and develops related ideas in coherent, sequential paragraphs.	
3	Essay presents a recognizable sequence of ideas though the thesis is omitted or paragraphing and structure are faulty.	
2	Essay relies on unrelated generalizations, vague argument, uncertain information.	
1	Essay is not coherent, showing little development of or relationship among ideas.	
	C. Conventions and Style	
5	Essay demonstrates facility with the conventions of academic writing and uses appropriate style.	
4	Essay generally shows facility with conventions of academic writing and generally uses appropriate style.	
3	Essay shows minor problems in flowing conventions of academic writing or using appropriate style.	
2	Essay presents problems in following conventions of academic writing or using appropriate style that hinder reader understanding or writer credibility.	
1	Essay is unacceptable in its style or ability to follow conventions.	
	D. Control of Mechanics	
5	Essay is generally free from errors in word choice and mechanics.	
4	Essay may have a few errors in word choice and mechanics.	
3	Essay has an accumulation of errors in word choice and mechanics.	
2	Essay is marred by numerous errors in word choice and mechanics.	
1	Essay has serious and persistent errors in word choice and mechanics.	

Appendix C: Used with permission

Speed Writing for Young EFL Learners

Tekka Chang, Junten Junior and Senior High School, Tokyo, Japan

Introduction

Teaching writing to young EFL learners can be quite challenging. Picture a grade 7 student in an EFL environment where only 2.3% of the population are foreigners and English is not one of the official languages of that country. He or she may not be very motivated to practice writing in English class. This is what I was facing when teaching grade 7 English in Japan. According to Japan's Ministry of Education, Culture, Science and Technology (2019, from hereon MEXT), starting from the 2011 changes to the Elementary School Japanese Curriculum Guidelines, students in grades 5 and 6 started learning English with a focus on listening and speaking, and reading and writing were primarily taught from grade 7. However, according to the 2019 changes, "what was learned through listening and speaking in elementary school was not transferring smoothly to reading and writing in junior high school" (MEXT, 2019, p.7) which starts at grade 7. In order to make the transition to reading and writing easier, I created a speed writing activity based on Kitahara's (2010) dictionary word look-up activity. In the reading activity, students race to look up new words in the dictionary at the back of the government-approved textbooks, and stand up when they found the word. Students can also help their classmates until every student is standing. I applied a similar concept to the speed writing activity as explained below.

Procedure

Step One

Prepare a handout that contains the new words and sentences with the key grammar point for the unit that you are going to teach. Below is an example of part of the key target sentences and new words in a grade 7 textbook that I have previously taught.

Table 1. Lesson 1 Part 1 Target Words and Sentences

Target Sentences	**Please write each sentence one time**
I am Tanaka Hana.	
I am fine.	
I play tennis.	
I play basketball.	
Target Words	**Please write each word 5 times**
happy	
sad	
angry	

Step Two

After teaching the new vocabulary for the day's lesson, assign students the unit vocabulary worksheet and ask them to take out a pencil to practice writing. Then create teams. One way is to assign teams by row and create a score sheet on the blackboard (see Table 1). For example, row 1 could be team 1, and row 2 would be team 2, and so on (if you have time, you can have the students create their own team name for that lesson too).

Table 2. Sample Scoresheet

Team 1	Team 2	Team 3	Team 4	Team 5
III	II	II	I	

I also make sure to change the way I make the teams every now and then. This way, a weaker student will not be picked on for slowing down his or her team.

Step Three

Inform students that you will read one of the new words. Students have to read the word that was said aloud while writing it. At this stage, it is important to tell the students that the writing must be clear enough to pass a test. Otherwise, the activity will get out of control. Students also need to read the words aloud so you can monitor your student's pronunciation. With large classes, this may not be possible but when I hear a word that many students are having trouble with, I make sure to practice as a class after the game is finished. After they have done this 5 times, they stand up. The first person to stand up in the row wins a point for their team and the first row to have all members standing will get a team point. This way, it is both an individual and team competition. In the past, if there was a really weak student in a team, he or she might be teased for being slow. Therefore, I added an individual competition element to prevent negative energy in the classroom. I usually start this activity by having students write single words by theme within the unit if possible followed by full sentences at the end.

Step Four

For key sentences that include the lesson's grammar point, read the sentence and students will only have to write the sentence once, but they still need to read each word out loud while they are writing. I keep full sentences at one repetition because I want students to work more on their sentence patterns in their textbook assignments which has meaningful context.

Step Five

Ask students to finish reading and writing each word 5 times for homework and 2 times for key sentences. The reasoning is that in deliberate learning, students need around 7 repetitions of the same word before they can learn it (Webb & Nation, 2017). If students write 5 times in class, 5 times for homework, and if they practice recitation of the readings that contain the new word, or key phrases in class, without stress, students will get enough reading repetitions to learn the new words for that lesson. I do not give any more writing at the beginning since my students will see the words again when doing their workbook that accompanies the text. I also create digital flashcards using the Quizlet app so students can increase the number of repetitions in another manner. This allows for more spaced repetitions in the following week during review practice.

Variation Activity

If you usually assign vocabulary and sentence writing as homework in a notebook, you could have students do this activity in their homework notebook. I ask my students to leave a line after each word so they can continue to do the remaining vocabulary writing at home next to what they have written in class.

Conclusion

Not all elementary school EFL programs start with a focus on listening and speaking like in Japan in the past. However, for many young learners, spelling practice can be a chore. The activity was especially effective for my male students that did not like to do homework. To them, this activity was like a race in their club activities. Furthermore, some students even re-wrote all of their vocabulary words. Some students said, "I want to re-write the words nicely so I can have beautiful writing." This simple reading and writing activity is a fun way to have students practice their writing while reading aloud. It also provides students with the required number of viewings of the new vocabulary necessary to learn the word in a fun manner. Since this activity is easy for students to do, each small achievement will boost students' confidence so that they can move on to more difficult reading and writing that will follow in the higher grades.

References

Kitahara, N. (2010). *Eigojyugyo no miki wo tsukuru hon* [A book to train strong foundations for English class]. Tokyo, Japan: Benesse Corp.

Ministry of Education, Culture, Science and Technology. (2019) Shogakko Gakushushido yoryou heisei 29 nen kokuji kaisetsu gaikokugo katsudo [Elementary School Japanese Curriculum Guidelines: Foreign Languages]. Retrieved from https://www.mext.go.jp/component/a_menu/education/micro_detail/__icsFiles/afieldfile/2019/03/18/1387017_011.pdf

Webb, S., & Nation, P. (2017). *How vocabulary is learned.* Oxford, United Kingdom: Oxford University Press.

About the Author

Tekka Chang is an English instructor at Junten Junior and Senior High School. He has a Master's in TESOL from Arizona State University and has taught English in Canada, Japan, and Taiwan. His research interests include extensive reading, second language acquisition, and pre-service teacher training.