A Survey of Primary School English Language Teachers' Language Competency in Ethiopia

Solomon Areaya *, Daniel Tefera ** and BelayTefera ***

Abstract

The purpose of this study was to assess English language teachers' English language competency. The teachers were teaching from grades 1 - 4 in six regions and one city administration of Ethiopia. A sample of 490 English teachers from 245 schools (two from each school) was selected. Data were collected using English language teachers' competency assessment (ENTCA) tool (measuring reading comprehension, writing, listening comprehension, and speaking). The major findings indicate that the overall performance of English teachers was relatively low, with an average percent-correct score of 40. Expressed in terms of performance levels, averaged across all ENTCA components, 52% of teachers performed at an insufficient level, 24% performed at a minimally acceptable level, 21% performed at a proficient level, and 3% performed at an advanced level. The targeted performance levels were proficient and above; the current percentage of teachers performing at these combined levels (24%) left a lot of room for improvement. Writing and speaking were the English performance skills that suffered the most among these grades 1-4 English teachers in public schools. The overall average percent-correct score for these two components was 20 and about 80% of teachers performed at this insufficient level. The least challenging English performance was listening comprehension, with an average percent-correct score of 63 and only 8% of teachers performing at the insufficient level. Hence, it was concluded that the majority of teachers were below the expected level of competency in English language skills and the expressive aspects of English language competency (writing and speaking) were even at critically low levels and, thus, require interventions for improvement.

Key Words: English competency, reading comprehension, writing skills, listening comprehension, and speaking skills,

*Associate Professor, Department of Science and Math Education, CEBS, AAU

** Assistant Professor, School of Psychology, CEBS, AAU

*** Professor of Psychology, School of Psychology, CEBS, AAU

Introduction

The issue of language teachers' competency is a critical factor in education. The success of education as a whole depends, among others, on the presence of those competent teachers who are prepared to stay in the teaching profession. This preparedness includes teachers' language knowledge and pedagogical skills in teaching the language as they are the key features of successful language teaching (Mayahi & Mayahi, 2014). As noted by scholars (e.g., Mata, 2014) language teachers should have the minimum competencies such as scientific knowledge, pedagogical competencies, values, and attitudes to teach a language.

Richards (2010) also described that for English language teachers to teach effectively, they should be able to comprehend texts accurately, maintain use of the target language in the classroom, give explanations and instructions in the target language, provide examples of words and grammatical structures, give accurate explanations, use appropriate classroom language, monitor his or her own speech and writing for accuracy , give correct feedback on learner language, and provide input at an appropriate level of difficulty. These are the minimum skills that English language teachers whose mother tongue is not English need to have in the target language in order to be able to teach effectively in English. Kelly et al. (2002) added to this point and described that the competencies should also include the personality attributes typical of a good language teacher, and how these affect classroom learning. Additionally, creating and maintaining a supportive classroom climate and the ways in which a teacher can create and sustain learner motivation, willingness to speak, and group motivation are qualities that language teaching requires.

Dinçer, Göksu, Takkaç, and Yazici (2013) analyzed a number of articles in language teaching and delineated four important qualities that language teachers particularly English language teachers should develop. According to these researchers, socio-affective skills (such as motivating students, sparing time for students when they ask for help, having positive attitudes towards students, responding to students' needs, and providing a stress-free classroom atmosphere), pedagogical knowledge, subject matter knowledge, and personal characteristics (such as being challenging and having reasonably high expectations, having sense of humor, being creative, being tolerant, patient, kind, sensible and open-minded, flexible, optimistic, and enthusiastic) are important qualities and language teachers should have a balanced combination of these qualities for better language teaching and learning.

Against this expectation, there is currently a perception in Ethiopia that language teachers are not as proficient as they should be in their use of English for teaching. For example, principals and supervisors who were participated in a study reported that most teachers of primary grades were not well trained to teach subjects in English (Solomon & Daniel, 2020). This situation may affect the quality of children's learning

to read and write in English which is an essential skill for learning all subjects beyond primary school as it can be seen in different kinds of assessments conducted in the country (e.g., NEAEA, 2020).

There are, accordingly, several reasons for assessing English teachers' competencies in Ethiopia now. First, the ongoing educational reform in Ethiopia as stipulated in the education roadmap has a significant focus on developing multilingual educational outcomes for students. English is one of the main languages that students are expected to transition to as a medium of instruction, including for higher education. Therefore, there is a need to establish the level of the knowledge and skill base of teachers for executing this assignment.

Second, teachers play a key role in improving students' learning outcomes. This is central to the education roadmap, informed by theories of teaching and learning. Research demonstrates that teachers cannot teach reading if they have not been taught how to do so. This is known as "the Peter Effect" (Binks-Cantrell, Washburn, Joshi, & Hougen, 2012). Further, teachers need to know how to teach reading to second or "later acquired" language readers in order to impart their own knowledge to their students.

Third, there is a strong theoretical reason for applying a "transition framework" when teaching a second or later acquired language. Learning to read a second or later acquired language is a significantly different process from learning to read a first language, primarily because the first language skills transfer to and impact (facilitate and interfere with) the process of acquiring second/later language reading skills (Koda, 2008; Nakamura, de Hoop, & Holla, 2018; READ M&E, 2019). Therefore, there is a need to assess whether English teachers have the necessary knowledge and understanding of the entire transition framework to support the development of English reading skills in their students.

Statement of the problem

As clearly indicated in the 1994 education and training policy of the country, the medium of instruction for secondary education and beyond in Ethiopia is English. This would entail that teaching English at lower grades would facilitate students' later learning and pave the way for their future carrier. In all five rounds of the national learning assessments (NLA) conducted at grades 4 and 8, students' achievement was not satisfactory based on policy expectations. In all these rounds, their achievement in composite score was below 50%. Similar results were observed in the sixth NLA conducted in 2020. The sixth assessment revealed that the mean percent scores for both grades in all core subjects were less than 50% of the passing mark set in the Ethiopian

education and training policy (NEAEA, 2020). The NEAEA report further indicates that students' performance on English is even worse for both grades compared to other courses. For example, the achievements of students in English in the sixth round were 30.75% and 33.64% for grade 4 and grade 8, respectively (NEAEA, 2020). These results are much lower than the preceding five rounds for both grade levels. This shows that, students are facing more difficulty in learning English. Thus, teaching English needs special attention, as English is the medium of instruction in most of second cycle primary schools, in all secondary schools and beyond in Ethiopia.

In this situation, the role of English language teachers is very important. English language teachers should be able to properly carryout their responsibilities to avert this condition. It is believed that poor quality of English language teaching leads to poor English language skill in students. The education sector would benefit from assessing English language teachers' competencies, because the results of the assessments would provide insight into the strengths and weaknesses of teachers in teaching English, which would inform policy decisions for improving in-service and pre-service teacher training of English language teachers.

Objective of the Study

This study aimed at exploring how well primary teachers are prepared to teach English as a subject in Ethiopia. Accordingly, the study attempted to answer the following questions:

- What is the overall performance of English language teachers on competency measure?

- What is English language teachers' level of performance across the components of the competency measure?

- Are there variations among English language teachers in the selected regions with respect to their competencies?

Methodology

The approach of this study in terms of content delineation and assessment was as follows: Content-wise, the approach of this study for assessing teachers' competencies was derived from both international best practices that conceptualize teacher competencies as well as a theoretical synthesis of the existing literature on competence in teaching reading (National Reading Panel, 2000), teacher competence (Roelofs &

Sanders, 2007), and reading acquisition in alpha-syllabic languages (Nag & Perfetti, 2014; Share & Daniels, 2015). English Teacher's Competency Assessment (ENTCA) was conceptualized and developed to measure teachers' competencies in teachers' own knowledge and skills in the language they teach that included: reading comprehension, vocabulary, grammar, writing, listening comprehension, and speaking.

In terms of assessment approach, the development of English Language Teachers' Competency Assessment (ENTCA) followed industry-standard procedures and began with the development of assessment frameworks. The typical assessment cycle, which was largely followed throughout the work on ENTCA, is depicted in Figure 1.
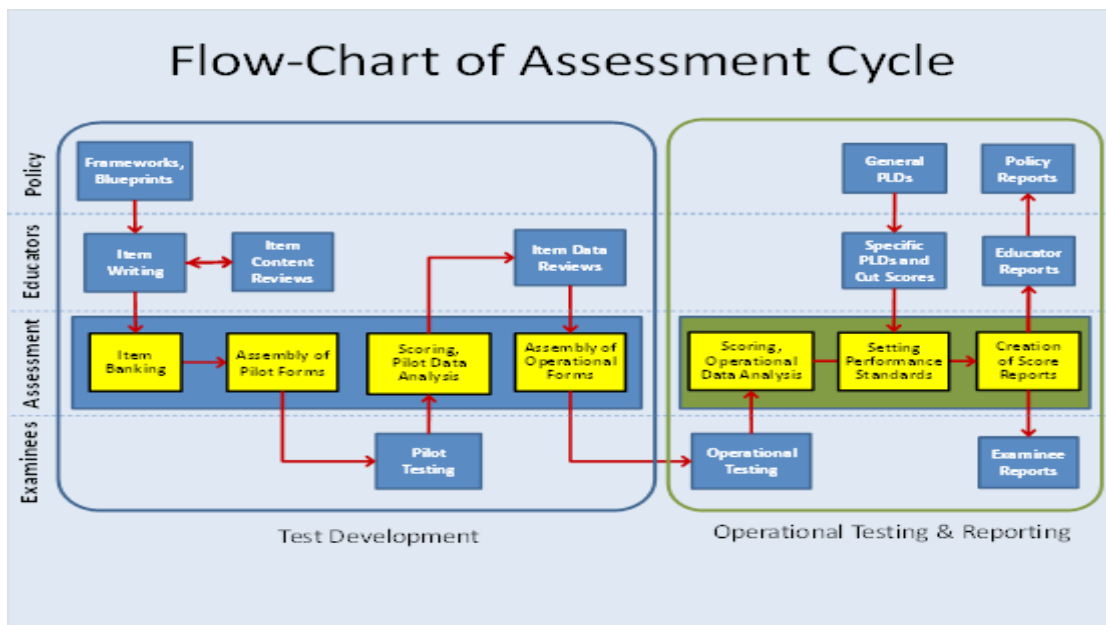

Figure 1: Flowchart of the Typical Assessment Cycle

The activities that were carried out in relation to each phase of test development process are described in the following paragraphs.

Item development and assembly of pilot forms

The item development workshop was conducted with the objective of developing teacher competency test items by organizing the item writing sessions involving education stakeholders such as MoE, NEAEA, and universities and colleges of teacher education. In total, 47 people (10 women and 37 men) took part in the workshop. Twenty-six tool developers (three women and 23 men) participated. The purpose of bringing all these stakeholders together was not necessarily for item writing but to partake on standard setting and the cut score determination process. Regarding the

actual item witting practice only those assessment experts from the NEAEA, MOE and CTE attended and accomplished the task after receiving intensive training by a high-level psychometrician.

Development of assessment frameworks and blueprints

The assessment frameworks for ENTCA focused on reading comprehension, vocabulary, grammar, and writing. They were designed and reviewed during the week of the item development workshop. Frameworks for the teacher assessments were designed to establish the measures of teacher competencies for teaching reading. Frameworks for the listening and speaking components of ENTCA were adopted from existing resources developed by a project funded by USAID Ethiopia.

Training of item writers

This workshop followed the procedural paradigm that started with providing extensive training in the principles of writing and reviewing quality items. Item writers for the ENTCA were recruited from among current and former staff of teacher education institutions, MOE language experts, curriculum experts, and test development methodologists. The criteria for selecting quality item writers were based on a combination of experience and achievement.

The training took several days and was combined with actual item-writing and peer-reviewing practice. The training focused on the appropriate alignment of items with assessment frameworks (content standards), coverage of various cognitive levels (especially higher order cognitive skills such as critical thinking and problem solving), gender fairness, and language quality. The study team ensured that items were competency-based and not textbook-based. Item writers were also trained to create scoring guidelines for all items (defining correct answers and the associated number of points), including the development of scoring rubrics for constructed response items (short and extended answers, and essays for writing). Item-writing training focused on the following major quality criteria.

- **Validity:** Alignment with curricula and standards

- **Reliability and discrimination:** Consistent scores in repeated applications and reliability in distinguishing between weak and strong test takers.

- **Objectivity and fairness:** Consistency between scores obtained by different scorers,

indicating item objectivity; and items do not favor or place at a disadvantage any group of students (e.g., on the basis of gender, ethnic background, rural–urban allocation, religion), indicating item fairness.

- **Efficiency:** Extracting as much information as possible information from a test item, in as little time as possible (e.g., a test item does not take 20 minutes to complete) and requiring as little space as possible (e.g., a single test item does not take a whole page).

- **Acceptability:** Item text and graphics use themes that are commonly acceptable from moral, political, and other perspectives.

The study team delivered training to item writers using a variety of materials, including a comprehensive *Item Writing Manual*, to ensure the creation of items that met industry standards of the highest technical quality. The manual included extensive examples for most of the quality principles.

Content review

The study team used rigorous procedures for reviewing item quality. During item writing, items were peer-reviewed and then submitted to a committee for further review to ensure that they met the criteria of aligning with the objective or standard being measured, and that they covered a prescribed range of cognitive complexity (according to the assessment framework). They were also reviewed for discriminatory language and content that may cause item bias, may be offensive to a specific group of examinees, or may inspire misconduct on the part of examinees. Bias and sensitivity are typically evaluated based on gender and other relevant examinee characteristics, as determined by MOE. Committee reviews were carried out by external groups of content experts whose qualifications were at least equivalent to those of item writers.

Following the week of item development, the study team worked with a smaller team of key subject specialists to assemble pilot test forms for the English teacher assessments. These pilot forms were built on corresponding blueprints and were equivalent to the operational forms for teachers. There were also linking items that were common across multiple forms. Two pilot forms were assembled for each test to provide a sufficient number of items for the construction of operational teacher assessments.

Pilot test administration, data entry, and data cleaning

The ENTCA tools were piloted in five regions. Regional and woreda-level experts were selected from each of the participating regions and zones to conduct the test

administration on groups of 50 English language teachers at their respective testing sites. A total of 335 English language teachers (95.7% of the targeted 350 teachers) from 262 schools participated in the tools piloting. All test administrators received orientation on how to administer the test, and the testing process was linear for all testing sites.

Table 1: Number of participants disaggregated by testing center, school number, and form types

Region/ zones	Testing site (city)	No. of schools	N. of teachers taking form 1	N. of teachers taking form 2	Total No. of teachers
Amhara	Bahir Dar	31	23	23	46
Oromia	Adama	44	22	22	44
Tigray	Mekele	45	25	24	49
Somali	Jigjiga	12	26	24	50
SNNPR					
Wolayta	Sodo	48	25	24	49
Sidama	Hawasa	35	25	24	49
Hadiya	Hosana	47	24	24	48
	Total	262	170	165	335

After administering all of the tests, examinee responses were digitized using both scanning technology and traditional procedures for manual data entry. The study team used widely accepted procedures based on scannable answer sheets to digitize test takers' responses on multiple-choice items.

For open-ended items, we used traditional human scoring procedures combined with manual data entry using low-cost data-entry software. We trained the scoring staff (mainly teacher educators who served as item developers) covering the scoring rubrics and the implementation of quality control procedures to ensure scorers were consistent, such as double scoring of items and the use of anchor documents.

Item analysis of pilot data

We carried out pilot test data analysis using its Data Management and Item Analysis System (DAMIAS) - a suite of SPSS macro programs that can be customized for data analysis of student and teacher tests. The system is designed to work in production mode, which means it can process many test forms and items, producing rich statistical outputs in a short time. The system also enables quality control at multiple stages of data analysis. The system's statistical output includes classical item statistics such as

item difficulty, item discrimination, option and score-point distribution analysis, and differential item functioning (DIF) analysis for a selected group of test takers. The DAMIAS system takes input from two files-the Test Map and the Original Data File which are processed using the following procedures.

- **Test map extraction:** This program extracts comprehensive information about the structure of the test (e.g., number of forms; number of items; their content, types, and position) and creates the files that will be used in further steps of analysis (e.g., the Item Bank File serves as a database where results from all item analysis will be automatically stored).

- **Data extraction:** This program carries out item scoring and computes test takers scores at the item, test, and sub-score levels, using both raw score and percent-correct score metrics. The SPSS Keyed Data File is created as a research database for further analyses of test taker- and item-level data.

- **Statistical analyses (classical test theory and item response theory):** Item analysis macro programs represent the heart of the system; they produce comprehensive item statistics, which are automatically stored in the Item Bank File and are ready for further use, such as exporting into specialized item banking platforms or producing item data review documents. Various graphs are produced and stored in designated folders in the form of image clips. Item response theory (IRT) item statistics are produced by external applications (e.g., Parscale, IRTPro, Conquest) and stored in the Items Bank File.

- **Quality control:** The system provides support for quality control checks, which are carried out at multiple points during data processing.

Item data review

Once the pilot data analysis was finalized, items were evaluated based on their content and statistics using item cards. These item cards are an alternative to an electronic item bank. They contain (a) the actual item, (b) information about the item content, (c) item statistics, and (d) differential item functioning (DIF), option, and score-point distribution graphs.

The study team prepared item cards for all pilot test items and then trained English language experts on their use during item data reviews. Each item was reviewed

based on industry-standard criteria for item construction, taking into consideration statistical performance. This enabled to make decisions regarding the status of an item- specifically, whether it would be (a) retained, (b) modified and re-piloted in the next round of testing, or (c) discarded.

Assembly of operational forms

The pool of items that was used to assemble operational forms came from the pilot administrations that were carried out following the item development workshop. Two full-length pilot forms had been administered, yielding a sufficiently large item pool for the assembly of operational forms. These items were also reviewed for their content and statistical quality, and the highest quality items were selected to assemble the ENTCA operational forms. The use of technology-enhanced tools, such as electronic item banking programs, was demonstrated and contemplated for future implementation. The quality of items was evaluated using both content and statistical criteria, in accordance with industry standards used for contemporary, large-scale assessment programs.

Sampling plan for operational administration

We implemented a combination of both purposive and random sampling. We also considered the existing COVID-19 situation, the proximity of schools to the assessment center, and security issues in the country as inclusion criteria for the selection of zones, woredas, and schools. All schools that took more than one day to reach the assessment center were excluded from the sample to minimize the risk of COVID-19 infection. All private schools and all zones and woredas with security problems were also excluded from the sample. Six regions and one city administration were purposively selected for inclusion. These regions and the city administration were believed to cover a sufficiently large geographic area in Ethiopia.

Within each of these regions and the city administration, 70 English teachers of Grades 1–4 were selected randomly to be included in the assessment. The number of teachers was decided based on power calculations, using parameters from a similar study. These power calculations determined that including 70 teachers per region and the city administration was sufficient to detect meaningful effects. Accordingly, we purposively as well as randomly selected 45 zones and 150 woredas from the selected regions and the city administration. We contacted the selected woredas and sub-cities through RSEBs, zone education departments (ZEDs), and the Addis Ababa City Administration Education Bureau to get lists of schools in the selected woredas and sub-cities. After the lists of schools were secured, sample schools were randomly selected.

The ENTCA sample contained a total of 245 schools. Two English teachers of Grades 1 – 4 from each school participated in the assessment, for a total of 490 teachers. Most first cycle primary schools (grades 1 - 4) in Ethiopia, particularly the rural schools, do not have more than three English teachers. An attempt was made to learn about the number of English teachers in the sampled schools before selecting individual teachers for the sample through communication with each region and the researchers were successful in receiving the maximum number of English teachers in each school. Accordingly, the researchers decided to select two teachers from each school for the purpose. The researchers believe this would represent nearly 40-50 percent of the English teachers available in each school.

From each of the randomly selected schools, two English language teachers were invited to participate in ENTCA. Accordingly, 490 English teachers teaching in grades 1 - 4 partake the assessment. The number of zones, the number of schools in which data were successfully collected, and the number of teachers assessed are presented in Table 2.

Table 2: Number of Zones, Schools and Teachers Included in ENTCA,by region

No.	Region &city administration	Number of zones	Number of schools	RCGV, Writing & Listening			Speaking		
				M	F	Total	M	F	Total
1	Addis Ababa	9	35	33	33	66	19	21	40
2	Oromia	12	35	37	32	69	25	15	40
3	Amhara	5	35	27	40	67	17	23	40
4	Benishangul	3	35	36	30	66	25	15	40
5	Tigray	6	35	31	38	69	23	17	40
6	SNNPR	6	35	45	24	69	26	14	40
7	Somali	2	35	58	4	62	33	4	37
Total		43	245	267	201	468	168	109	277

Setting Performance Standards for ENTCA

Setting performance standards is a process of developing *conceptual* and *operational* definitions of knowledge, skills, and competencies that teachers should possess at a specific mastery level. Traditional content standards specify *what* teachers are expected to know and be able to do in a given area of expertise. Performance standards specify

how much teachers are expected to know and be able to do at a particular performance level (e.g., basic, proficient, or advanced).

Conceptual and operational definitions of performance standards are a necessary component of a standards-based assessment system for evaluating teachers' performance regarding specified achievement criteria. Incorporating performance standards into national assessment systems as descriptions of national learning targets substantially improves the interpretability of assessment results and provides a meaningful framework for monitoring educational progress at the institutional, regional, and national levels. There are *two stages* of setting performance standards: setting performance levels, and setting cut scores between the performance levels.

Development of performance-level descriptors

Setting performance levels is a procedure for conceptualizing performance levels for the evaluation of teacher performance-specifically, deciding on the number and purpose of performance levels, choosing their labels, and developing general and specific descriptions of each level (performance-level descriptors [PLDs]; e.g., what primary English teachers should know and be able to do in order to be considered "proficient"). This procedure is based on information drawn from official documentation on national learning standards and curricula. The procedure for setting performance levels typically employs a focus group discussion method to solicit the opinions and judgments of field experts (e.g., teacher educators, supervisors, curriculum experts, MOE officials). Figure 2 summarizes the process for setting conceptual definitions of performance levels.
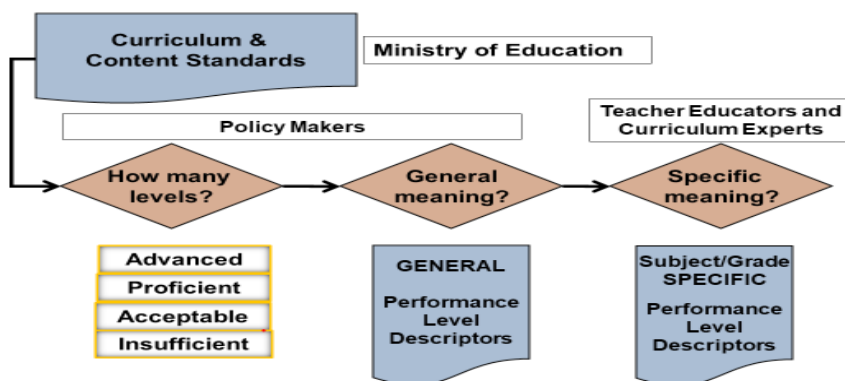

Figure 2. Flow of activities for setting performance levels

Earlier before the undertaking of the current study a workshop was convened for developing conceptual definitions of Performance Levels and PLDs for MT teachers in seven Ethiopian languages, in December 2019 in Addis Ababa. There were 58

A Survey of Primary School English Language …

participants, including professors and instructors from teacher education institutions, supervisors, and curriculum experts. The general PLDs developed during that workshop-which are designed to support the evaluation of teacher performance in any subject in primary education in Ethiopia-were used as a starting point for developing specific PLDs for ENTCA. We adapted specific PLDs developed for MT teachers into PLDs that were applicable to ENTCA. Forms for the estimation of cut scores were also designed.

Estimating cut scores

Setting cut scores is a procedure for establishing score ranges on an operational test scale to classify teacher performance into established levels. All methods for setting cut scores rely on obtaining judgments from subject-matter experts and then mapping conceptual definitions of performance levels onto score ranges on actual assessment instruments. Those score ranges represent operational definitions of performance levels, and their meaning is maintained across different test forms (sets) within and across administration years through test-equating procedures. Figure 3 summarizes the process for setting cut scores.

Due to a lack of resources and COVID-19 restrictions, a simplified procedure was followed to set cut scores for ENTCA. Judgments were obtained from five experienced English language experts, who also participated in test development activities for ENTCA. The experts' judgments were obtained through e-mail communication, presenting the detailed instructions for estimating cut scores, estimation forms, and the actual test booklets and recording scripts (for listening and speaking). As all the experts had participated in test development steps for ENTCA, and most had also participated in setting performance standards for MTTCA, they reported that the estimation of cut scores was not a challenging process. The study team successfully collected data and created the cut scores for the ENTCA components, which were then applied to calculations of performance-level scores.
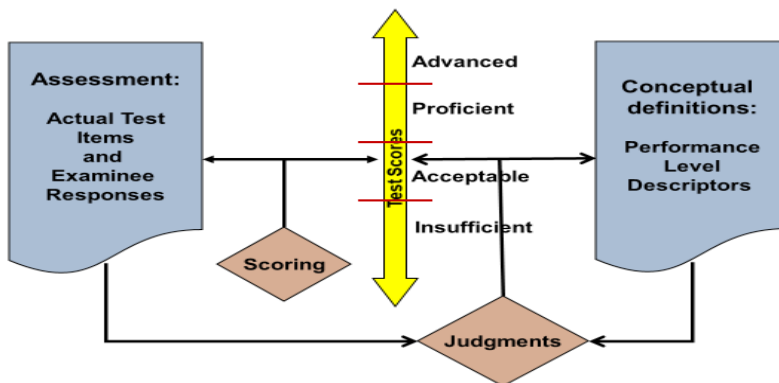
Figure 3. Flow of Activities for Setting Cut Scores

Data analysis

The data analysis was managed by a senior psychometrician and statisticians using the following steps. First, the SPSS-based DAMIAS was used to manage data, conduct item analysis under classical test theory (CTT) models, and produce examinee scores necessary for further analyses. For the ENTCA component that was not piloted (listening comprehension), this stage enabled the quality of items to be scrutinized, which eventually led to the removal of one item that did not meet statistical quality requirements.

Then teacher performance on each of the ENTCA components was analyzed across regions, for urban versus rural locations, and by teacher gender. The statistical method used in these analyses was analysis of variance (ANOVA). In addition to statistical significance, information about effect size was provided using Cohen's *d*.

Results and Discussion

This section discusses the results of the ENTCA assessment. We first present scores as percent correct, relative to the maximum possible score on ENTCA, for both the total score and sub-scores. While the percent-correct score is straightforward, transparent, and easy to understand, it does not convey sufficient information about the value of the result because it is affected by the difficulty of the test (easy tests generate high scores, and hard tests generate low scores). For this reason, we also report scores based on performance standards. This addresses the limitation of percent-correct scores by categorizing teacher achievement into meaningful and readily interpretable

A Survey of Primary School English Language …

performance levels. These levels were determined based on the common conceptual definitions of teacher performance standards, and their operational definitions (in terms of cut scores on instruments) were developed taking into consideration nature of specific components of ENTCA. The measure we used is the percentage of teachers whose English test achievement fell within each performance level (insufficient, acceptable, proficient, and advanced), according to the established standards.

In the remainder of this section, teacher achievement on ENTCA is presented as average percent-correct scores and performance levels for each main component: reading comprehension, writing, listening and speaking. Results are presented for the total skill score, as well as sub-skill scores.

Teachers' language and literacy knowledge and skills

ENTCA is broken down into five main components encompassing language and literacy skills. In this section, we present the results for teachers' English language and literacy knowledge and skills, focusing on three sub-components: reading comprehension, writing, and listening and speaking. Results for each sub-component are presented as total component scores and sub-scores. All total scores are also disaggregated by region, teacher sex, and urbanicity.

Reading comprehension, vocabulary, and grammar

The reading component of ENTCA 2020 consisted of tasks related to reading comprehension, vocabulary, and grammar. We present the averages of the total percent-correct scores for the overall reading component (using tables and figures).

Reading Percent-Correct Scores by Sex

We examined reading comprehension percent-correct scores by teachers' sex and by region. The results of this analysis are presented in Table 4. Overall, we found a significant difference between male and female teachers, with a relatively small effect size ($p=0.003; d=0.28$).

Table 3. Descriptive Statistics for Reading Percent-Correct Score by Sex

		Mean	Standard Deviation	N	P-value	Cohen's d
Total	Male	50.67	20.649	252	0.003	0.28
	Female	44.74	21.915	205		

Reading Percent-Correct Scores by Urbanicity

We also examined the reading comprehension percent-correct scores by urbanicity (i.e., whether the teacher was in a rural or urban area). Overall, we found that teachers in urban areas significantly outperformed their counterparts in rural areas (*p =0.000; d=1.11*) by a meaningfully large differential (63.91 compared with 42.50).

Table 4. Descriptive Statistics for Reading Percent-Correct Score by Urbanicity

		Mean	Standard Deviation	N	P-value	Cohen's *d*
Total	Urban	63.91	22.076	115	0.000	1.11
	Rural	42.50	18.439	351		

Total Writing Percent-Correct Scores by Sex

When examining differences in writing scores by teachers' sex, we found that males significantly outperformed females (*p=0.000*), though there was a relatively small Cohen's *d* associated with that difference (*d=0.34*).

Table 5. Descriptive Statistics for Total Writing Percent-Correct Score by Sex

Region		Mean	Standard Deviation	N	P-value	Cohen's *d*
Total	Male	22.47	23.706	253	0.000	0.34
	Female	14.81	21.905	206		

Total Writing Percent-Correct Scores by Urbanicity

Overall, we found large discrepancies in scores between urban and rural areas, with teachers in urban areas scoring 37.76 on average, and teachers in rural areas scoring 12.95 on average. This difference was statistically significant and resulted in a large Cohen's *d* (*p=0.000; d=1.20*).

Table 6. Descriptive Statistics for Total Writing Percent-Correct Score by Urbanicity

Region		Mean	Standard Deviation	N	P-value	Cohen's *d*
Total	Urban	37.76	28.205	115	0.000	1.20
	Rural	12.95	17.717	353		

Listening and speaking

Listening and speaking skills were measured through a listening comprehension

A Survey of Primary School English Language …

assessment and an oral conversational assessment. Recognizing that these two components measure different (albeit similar) constructs, we present the results from each assessment in this section.

Table 7. Descriptive Statistics for Total Listening Percent-Correct Score

Region	Mean	Standard Deviation	N
Total	63.01	16.745	466

Similarly, Table 8 shows the scores for speaking. On average, across all regions, teachers scored 63.01 on the listening assessment and 21.41 on the speaking assessment.

Table 8. Descriptive Statistics for Total Speaking Percent-Correct Score

Region	Mean	Standard Deviation	N
Total	21.41	20.870	252

In the full sample, we found no statistically significant differences in average listening or speaking scores for male and female teachers.

Table 9. Descriptive Statistics for Total Listening and Speaking Percent-Correct Score by Sex

	Region		Mean	Standard Deviation	N	P-value
Listening	Total	Male	64.31	16.711	251	0.110
		Female	61.81	16.551	206	
Speaking	Total	Male	22.22	21.457	140	0.372
		Female	19.88	19.287	110	

As depicted in the table above, the differences in the average performance of male and female English teachers on both listening and speaking tests were not statistically significant than female teachers.

Total Listening and Speaking Percent-Correct Scores by Urbanicity

We examined listening and speaking scores by urbanicity as well. Total average scores for each construct showed statistically significant differences between teachers in urban areas and teachers in rural areas. The average listening score in urban areas was 75.41, compared with 58.95 in rural areas ($p=0.000; d=1.09$). The average speaking score in urban areas was 37.35, compared with 16.10 in rural areas ($p=0.000; d=1.14$). The differences based on urbanicity were driven by teachers in Oromia for both constructs, as well as teachers in Amhara for listening.

Table 10. Descriptive Statistics for Total Listening and Speaking Percent-Correct Score by Urbanicity

	Region		Mean	Standard Deviation	N	P-value	Cohen's d
Listening	Total	Unban	75.41	13.925	115	0.000	1.09
		Rural	58.95	15.567	351		
Speaking	Total	Unban	37.35	25.498	63	0.000	1.14
		Rural	16.10	15.923	189		

Teachers in urban schools performed better than those in rural schools in both listening and speaking tests. Teaching requires both skills and teachers who are better in speaking skill are advantageous in terms of delivering the lesson.

Listening and Speaking Performance Levels

This section presents teachers' performance levels for both listening (Figure 4) and speaking (Figure 5). Teachers across regions included in assessment appeared to have a decent grasp of English listening skills; the majority fell within or above proficient levels, except for teachers in Somali. Conversely, most teachers in regions included in assessment had insufficient speaking skills based on their assessment scores.
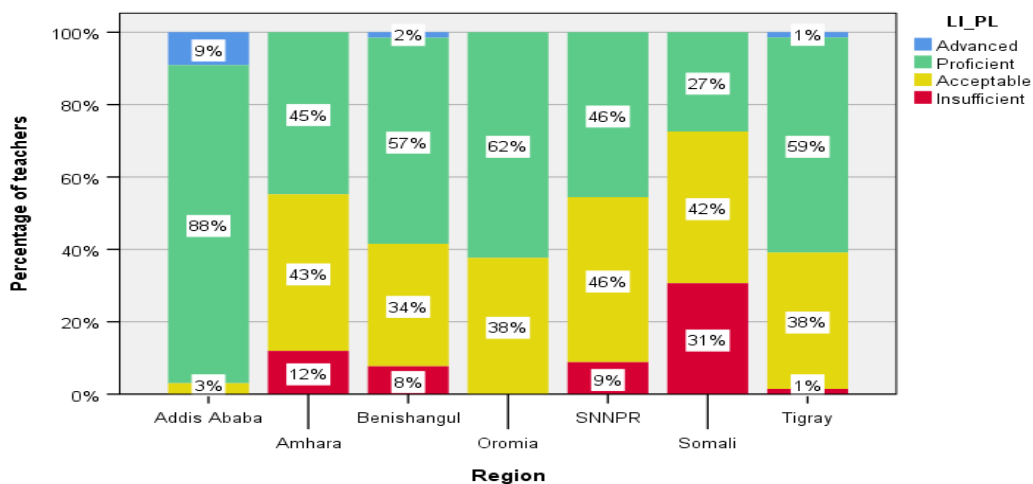

Figure 4. Percent of Teachers Reaching Total Listening Performance Levels by Region

A Survey of Primary School English Language …

As Figure 4 shows, relatively large (9%) of teachers from Addis Ababa schools and none from schools selected from Amhara, Oromia, SNNPR, and Somali regional states reached advanced level in their listening skill. On the other hand, no teachers from Addis Ababa and Oromia and only 1% of teachers from Tigray region performed at the insufficient level in the listening assessment test.
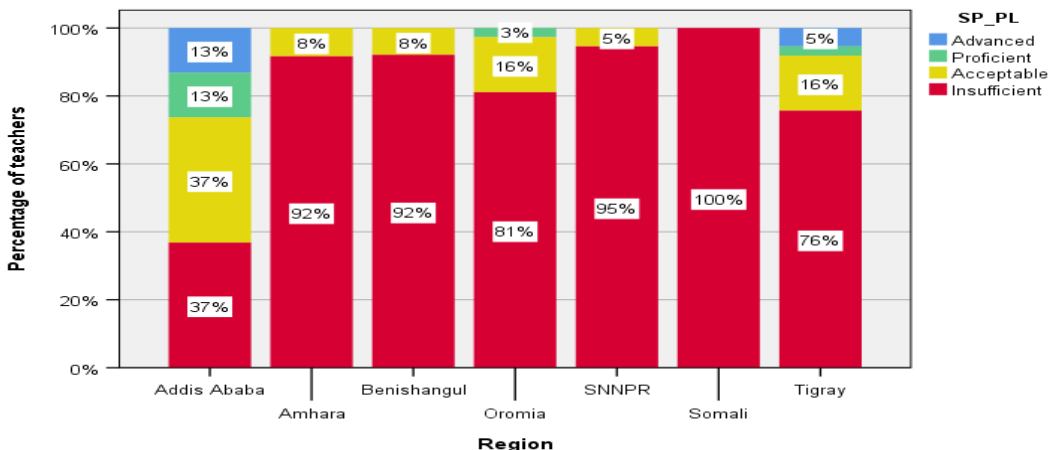

Figure 5. Percent of Teachers Reaching Total Speaking Performance Levels by Region

As was described earlier, grades 1-4 English language teachers' performance on speaking task is generally poor. When seen across regions, the data show that all teachers from Somali region and more than 90% of teachers from Amhara, Benishangul-Gumuz, and SNNP regional states performed at the insufficient level. On the other hand, 13% and 5% teachers from Addis Ababa and Tigray, respectively, performed at advanced level on speaking task.

Conclusions

This concluding section summarizes the key findings, based on ENTCA data collected from English teachers in seven Ethiopian regions. The results of ENTCA offered critical evidence to inform discussions and decision-making regarding policies, professional development, strategies, and interventions to improve both English language teachers' competencies and students' performance in English in the primary grades, and as a result, student learning outcomes across all stages of education in Ethiopia.

The study's main finding was that there was large variance in all components of

English language teachers' competencies across the selected 'regions. Addis Ababa city administration consistently stood out as the highest performing region, exhibiting much higher scores than the other regions. Somali Region was at the bottom end of the distribution, consistently delivering the lowest teacher competency scores.

Gender differences in the teachers' competencies were significant in two ENTCA components (reading and writing), with male teachers performing better than female teachers. There were some variations in gender differences across regions, with Amhara, Benishangul-Gumuz, and Oromia all showing significant gender differences in performance in favor of males in virtually all ENTCA components.

Differences in teacher performance between rural and urban areas were dramatic, with teachers in urban areas outperforming teachers in rural areas (with substantial effect sizes) in all components of ENTCA.

When looking at teacher performance in different components of ENTCA in relation to performance standards set by local experts, teacher performances were critically low in the expressive aspects of the English language (writing and speaking), with around 80% of teachers performing at the insufficient level for these two components. Teacher suggests that there is considerable opportunity for improvement, as about 40% of teachers performed at the insufficient level. The listening comprehension component appeared relatively easy for teachers, with only 8% performing at the insufficient level and 57% performing at the proficient level or above.

The results from this study suggest the need for further investigation of the factors that contribute to large differences in English language teachers' performance across different regions in, between rural and urban areas, and between male and female teachers in some regions. The ENTCA tool can play an important role in identifying the strengths and weaknesses of the teacher cadre in targeted regions, which can in turn inform actions to improve specific aspects of English teacher competencies. For example, the results of this study suggest that competencies in the expressive aspects of English language (writing and speaking) are at critically low levels and require action.

References

Binks-Cantrell, E., Washburn, E. K., Joshi, M. R., & Hougen, M. (2012). Peter Effect in the preparation of reading teachers. *Scientific Studies of Reading, 16*(6), 526–536.

Dinçer, A., Göksu, A., Takkaç, A., & Yazici, M. (2013). Common Characteristics of an Effective English Language Teacher. *The International Journal of Educational Researchers*, 4 (3), 1-8. ISSN: 1308-9501 http://ijer.eab.org.tr

Kelly, M., Grenfell, M., Gallagher-Brett, A., Jones, D., Richard, L., & Hilmarsson-Dunn, A. (2002). *The training of teachers of a foreign language: Developments in Europe*. A Report to the European Commission, Directorate General for Education and Culture

Koda, K. (2008). Impacts of prior literacy experience on second language learning to read. In K. Koda & A. M. Zehler (Eds.), *Learning to read across languages: Cross-linguistic relationships in first- and second language literacy development* (pp. 68–96). New York, NY: Routledge.

Mata, L. (2014). Pedagogical Competencies for Mother-Tongue Teacher Education. *Educational Sciences: Theory & Practice*, *14*(1); 341-352. DOI: 10.12738/estp.2014.1.1723

Mayahi, N. & Mayahi, F. (2014). "Isn't it our fault?" Teachers Language Knowledge and Skills. *Procedia - Social and Behavioral Sciences*, *98*, 1119 – 1127. Retrieved from www.sciencedirect.com

Nag, S., &Perfetti, C. A. (2014). Reading and writing: Insights from alpha syllabaries of South and Southeast Asia. *Writing Systems Research, 6*(1), 1–9.

Nakamura, P. R., de Hoop, T., & Holla, C. U. (2018). Language and the learning crisis: Evidence of transfer threshold mechanisms in multilingual reading in South India. *The Journal of Development Studies*, DOI: 10.1080/00220388.2018.1493192

National Educational Assessment and Examinations Agency [NEAEA] (2020). *Ethiopian Sixth National Learning Assessment Study of Grade Four and Eight Students*. Addis Ababa, Ethiopia.

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. National Institutes of Health, NIH Publication No. 00-4769. Retrieved from https://www.nichd.nih.gov/publications/pubs/nrp/smallbook

READ M&E, (2019): *EGRA 2018 Endline Report*.

A Survey of Primary School English Language …

Richards, J. C. (2010). Competence and Performance in Language Teaching. *RELC Journal, 42* (2), 101-122. https://doi.org/10.1177/0033688210372953

Roelofs, E., & Sanders, P. (2007). Towards a framework for teaching competence. *European Journal of Vocational Training, 40,* 123–139.

Share, D. L., & Daniels, P. T. (2015). Aksharas, alphasyllabaries, abugidas, alphabets, and orthographic depth: Reflections on Rimzhim, Katz, & Fowler (2014). Writing Systems Research, 1–15.

Solomon Areaya & Daniel Tefera (2020). Stakeholders' Perceptions Related to Transition from Mother Tongue to English Medium of Instruction. *The Ethiopian Journal of Education, XXXX* (1), 93-132.