

Observational Technique as an Assessment Agenda: Teacher and Student Perceptions

Judith Ojung'a

University of Eastern Africa, Baraton, P.O. Box 2500, Eldoret
judithojunga@yahoo.com

Hulda Amenya

University of Eastern Africa, Baraton, P.O. Box 2500, Eldoret
achenya@yahoo.com

Elizabeth Role

University of Eastern Africa, Baraton, P.O. Box 2500, Eldoret
bethrole@gmail.com

Lazarus Ndiku Makewa

University of Eastern Africa, Baraton, P.O. Box 2500, Eldoret
ndikul@gmail.com

Abstract

This study looked at teachers' and students' views on the use and usefulness of observational techniques in the teaching and learning of English. Twenty nine (29) teachers and 1030 students participated. Cronbach's Alpha Coefficient was used to estimate the internal consistency of items in the questionnaire and a coefficient above 0.6 was accepted as reliable. The Mann-Whitney test was used to determine the difference in the perception of students and teachers on the teachers' use and usefulness of observational techniques. The Kruskal-Wallis (K-W) test was used to determine the difference in perception of teachers on the use and usefulness of observational techniques. Scoring rubrics, anecdotal records and portfolios have similar impact when used in the evaluation of performance, while checklists, rating scales, and usefulness of observational techniques have different impact when used in the evaluation of English performance. Students in the girls' schools find observational techniques more useful than the students in boys' and mixed schools, but the teachers were in agreement that the techniques are beneficial to all schools irrespective of their types. Teachers and students had varied opinions on the use and usefulness of observational techniques in Provincial and District schools. The students' accepted that only scoring rubrics could be used in both Provincial and District schools while the teachers agreed that all the observational techniques were useful in evaluating performance in both Provincial and District schools.

Keywords: Observational techniques, scoring rubrics, anecdotal records, rating scales, Kenya

1. INTRODUCTION

It would be necessary to understand the meaning of assessment in general before the issue of use of observational techniques is looked at. The term 'assessment' (in education) according to Kellaghan and Greany (2001) may be used to refer to any procedure or activity that is designed to collect information about the knowledge, attitudes, or skills of a learner or group of learners. A more detailed definition is that assessment is "the process of obtaining information that is used to make educational decisions about students, to give feedback to the student about his or her progress, strengths and weaknesses, to judge instructional effectiveness and curricular adequacy, and to inform policy". Assessment, according to Petty (2004), measures the breadth and depth of learning. Assessment has traditionally been associated with the appraisal of individuals.

Assessment has been criticized as being inaccurate and unreliable (Petty, 2004). It is also true that assessment results are notoriously poor at predicting future performance. And yet society and teachers are unable to manage without it. In the right hands, assessment can inspire, motivate and provide feedback which is essential for targeting prompt corrective help, but it can also lead us to ignore what cannot easily be measured. Danielson (2002) says that assessment methodologies must, however, be appropriate to the type of goals being assessed. Although short-answer or multiple-choice questions might be suitable for goals related to knowledge, essay questions or oral presentations are more appropriate for measuring communication or reasoning skills. Some dispositional goals such as persistence or aesthetic appreciation – are virtually impossible to assess formally; in these cases anecdotal accounts may have to suffice. Teacher observations and pupil observations combined together help present a complete picture of pupils' adjustments, interests and attitudes (Gronlund & Linn, 1990). Assessment should never be used as a punishment, but should be used to validate learning. According to Simmons (2009), secondary English is definitely not one of the easier subjects to assess because unlike other subjects, it allows for multiple answers, all differing. An assessment of writing, for example, provides an indication of how well students communicate in written language. Teachers can determine student progress in writing through direct measures of functional literacy, such as writing samples, process writing, and dialogue

journals (Pierce & O'Malley, 2009). Students should have a sufficiently clear understanding of what to expect from a test to confidently say, "Ask me anything – I'm ready!" (Danielson, 2002).

Assessment should not force students to compete against one another; any competition should be between students and their own prior performance. Danielson (2002) says it is not only the "top 5" students who receive an "A"; any student who exceeds the standard established for an "A" should receive one. In other words, students are not trying to be better than others in the class but to be as good as they can. This orientation sets up an entirely different climate within the class.

There are many purposes of assessment some being only tangential to the interests of school staff for detailed instructional planning. The main use of assessment for teachers is the ongoing or formative assessment, which is used throughout the course to form judgments on whether, and to what extent learning has been successful and also to pin point difficulties so that remedial action can be taken. Danielson (2002) says this is not trivial information and can in fact provide schools, Districts or states with feedback about the effectiveness of their programs. Furthermore, assessment results are valuable not only as one-short data, but can also reveal much about trends and patterns, allowing educators to set improvement goals. Moreover, assessment and evaluation practices must be well planned, tied to the curriculum and capable of meeting student needs (Papert, 2004).

It is because of the above reasons that the researchers sought to assess the use and usefulness of observational techniques to evaluate performance in English. Both teachers and students were expected to evaluate their performance since assessment is often used for analysis of instruction and to determine strengths and weaknesses as well as aspects of instruction that need emphasis.

In the article, *English Teachers' Perception of Literacy assessment in the first Year of secondary school*, Moni et al. (1999) in their study found that when year 7 and year 8 teachers explicitly teach principles of assessment derived from the syllabus to their students, involve students in the assessment process, share their assessment practices, and liaise closely with each other, clarity and consistency of literacy assessment programs in primary and secondary schools are enhanced. The authors propose that achieving these goals is crucial to ensuring that literacy assessment in secondary school is a positive experience for incoming students. In Australia for example, Moni et al. (1999) found that the move to secondary school is a period of great educational change for many children where the majority attend state schools. Changes the children experience at this time include differences in school organization, classroom climate, instructional practices, and peer and teacher relations. Meyer (2009) also, highlights the experiences of many students on summative assessment as the dominant mode of assessment in most South African schools which result in students entering university ignorant of good formative feedback. They are therefore often unable to recognize its value and may even be traumatized by the presence of much ink on their pages. In order to achieve clarity and consistency, more information needs to be made available in the educational community about teachers' perception of their literacy assessment programs in the first year of secondary school.

1.2 Performance in English in High Schools

Baines (2008) reported that only 57% of 2005 tested high school graduates in the United States of America were ready for college-level reading. In 2002, 28% of 4th graders, 31% of 8th graders and 24% of 12th graders performed on or above the proficient level in writing. Only 2% of students in each grade performed at advanced level in writing (National Centre for Education Structures, 2005)

With the surge in use of non print media, reading books has been on the decline. Research indicates that the time students spend reading is getting crowded out by time spent with electronic media-particularly, video games and of course, television. In response to the decline of reading (especially during adolescence) some recent initiatives such as the federal governments high school initiative seek to enhance achievement among high school students through a reemphasis on reading and writing. Another alarming statistic surfaced in literacy trends. The national assessment of adult literacy (Kutner et al., 2007) found that the number of 16-to 18- year-olds considered literate actually has decreased since 1992. Another finding was that literacy of college graduates nosedived during the same period. The percentage of college graduates with proficient literacy decreased from 40% in 1992 to 31% in 2003 (National Assessment of adult literacy, 2005)

In the year 2006 Kenya Certificate of Secondary Education report, the Ministry of Education released a report on the candidate's performance in the three papers offered in English as follows:

In Paper 1, the mean score was 30.71 out of a maximum score of 60;

In paper 2, it was 29.88 out of a maximum score of 80; and

In paper 3, the mean score was 18.93 out of a maximum of 60 (KNEC, 2007)

In functional writing candidates failed to get the format and tone of language right. In cloze test, there was inability to grasp the context, grammatical inaptitude and limitations of vocabulary. In the oral narrative, the candidates gave stock answers to questions. Poetry continues to elicit poor performance. In the comprehension passages, candidates lifted sentences from the passage which did not respond to the question adequately. Candidates also displayed little knowledge of the text. Majority of the candidates also were not at home with

some of the terms and the implications the terms evoked and so wrote compositions about anything and tugged the required ending on it (KNEC, 2007). The performance is not better in Nandi Central District.

Something ought to be done to improve on students' performance. Moni et al. (1999) say that at a time of great public debate about the decline or otherwise of literacy standards, understanding more about literacy assessment in different sectors of schooling enables teachers not only to teach and assess literacy consistently and coherently, but also to be confident and competent advocates of their practices. The above situation therefore, indicates, there is need for explicit focus on formative evaluation using observational techniques and its usefulness for students. It is hoped that the use of observational techniques to evaluate performance in English will boost students' performance.

1.3 How Teachers should Formatively Assess using Observational Techniques

In the end teachers must work their own way to assist English learners but Petty (2004) suggests a few very effective strategies; assessment performances, mastery tests, and self peer or spoof assessment. Other useful strategies for assessing and keeping records in the classroom are portfolios (Hempek, 2009). Other effective evaluation strategies suggested are anecdotal records, celebration of learning, exit cards, graphic organizers, journals, oral presentation, peer assessment, project- based learning, rubrics, simulation, Checklists, Conferencing, Interviewing, Observation, Performance events/exhibitions, and Self assessment/ reflection. The above ensure well rounded grading practices. They are classroom based assessments for continuous programs. Using them allows assessment practice in combination for a better look at a students' progress with standards. You will no doubt use traditional multiple choice, fill in the blank, true or false type assessment but for those assessments that aren't nearly as straight forward, other methods are necessary.

Papert (2004) says that assessment tools are needed to guide students and teachers in setting appropriate learning goals because, we should be ensuring what students can do with knowledge not how many right answers they can give to questions. This study investigated on the use and usefulness of the following observational strategies: use of scoring rubrics, anecdotal records, portfolios, checklists, and rating scales. These strategies provide useful means to evaluate and record the assessment of student's on-going performance in English, consequently ensuring that the candidates presented for Kenya certificate of secondary examinations (KCSE) are well prepared. Yi (2012) emphasizes that formative evaluation is a method of evaluating a program while the programme activities are forming or happening. Yi further notes that formative evaluation focuses on the process during which problems are timely spotted, corrected and adjustments made so that the intended goal can be accomplished.

1.4 Scoring Rubrics

Papert (2004) defines a rubric as "a road map, telling students and teachers where to begin, where they're going and how to get there" (p.1). Rubrics are scoring guides or sets of expectations or criteria used to assess student level of understanding and allow students to know the expectations and what they need to do in order to be learning at a higher level. Some individuals believe rubrics help teachers and students focus on what is valued in a subject, activity, topic etc., (Airasian, 2000). Hempeck (2009) further adds that rubrics are a great way to assess a students' progress with standards in a detailed manner. Hempeck gives the following purposes of rubrics:-

1. For providing answers to assignments that require more than a simple selection of the correct answer or filling in the blanks.
2. Rubrics provide several focal points when grading subjective work.
3. They also provide the teacher with a tool to ensure that grading is fair and consistent for all students.
4. Rubrics provide a primary focus for otherwise ambiguous assignments.

Scoring rubrics are necessary for all the performance assessment methods such as projects and portfolios and for scoring essays and show- the- work problems. In stage one of the crafting process, achievement dimensions and a scale of progress for each dimension is identified from very low to very high progress. To craft a scoring rubric, there is need to refine these descriptions of performance levels to be sure they are clear. You may associate each level with a numerical value or alternatively associate each level with a qualitative description such as novice, apprentice, proficient and distinguished. It is necessary to describe the characteristics of a students' performance that distinguish one achievement level from another, because these descriptions anchor the scale at each level.

1.5 Anecdotal Records

This is where you observe the performance and write a description of what the students did. Pierce and O'Malley (2009) say anecdotal records are notes based on teacher observation of how students learn. They are a form of ongoing assessment of observations of student(s) in the classroom. They usually include the date, time, events, setting, students' name and teachers' name. Hempeck (2009) says they are a great way to document student behaviors and academic progress over time. It involves a teacher taking brief notes on a student's interactions within the classroom with subject matter and peers.

Taking observational notes allows the teacher to record a wide range of authentic experiences and even unintended outcomes of literacy development (Boyd-Batstone , 2004). Further, these notes are used to record objective and subjective information as well as affective information, such as levels of engagement, curiosity, and motivational factors. The American Association of School Administrators note “An anecdotal record is a written record kept in a positive tone of a child’s social, emotional, physical, aesthetic, and cognitive development (Airasian, 2000). Regular anecdotal recordings offer glimpses of progress and patterns of behavior not necessarily captured by other means of assessment.

By keeping such notes, the teacher has a point of reference for planning for instruction as well as an additional tool for communication to parents of their child’s progress. Watson and Steege, (2003) describe performance-based recording, which involves the rating of observed behaviors according to a predetermined scale. This type of recording is appropriate for instances when the behavior is observed firsthand but it is not feasible to record data at that moment. Instead, a rating of the behavior’s frequency, duration, or latency is made when the observer is able to record. Anecdotal records, therefore, facilitate assessment conversations as educational professionals describe their observations of student learning and consider ways to develop appropriate strategies to build on strengths and address academic needs. The more focused the observational records, the more helpful they can be in making daily decisions about instructional approaches.

The open-ended nature of the anecdotal record allows teachers to describe in rich detail each student’s development in literacy, and to integrate these observations with other available information, then identify instructional approaches that may be appropriate (Boyd-Batstone , 2004). They can be used to determine a student’s integration of reading and writing process (Pierce & O’Malley, (2009). They are also an essential component in the development and interpretation of student portfolios. In addition, permanent products (math worksheets, handwriting sample, essays or compositions) recording, which involves the inspection of tangible or “permanent materials completed by the students, can provide important information regarding the frequency, intensity, and severity of a behavior (Brown-Chidsey, 2008). Further, Brown-Chidsey observed that observational records can be used to assess the language proficiency of English learners by suggesting that for students with spelling difficulties, permanent products of the students recent spelling tests may provide information regarding typical or common spelling mistakes..

Consequently, being a teacher calls for skilled techniques in observing children, recording, and managing authentic assessment data (Boyd-Batstone, 2004). Recording observational data explicitly depends on the human expert, the kid watcher, and the sensitive observer. In other words, the one closest to the classroom experience is in a unique position to see and communicate a reliable and valid instructional perspective of the child”.

One method of keeping anecdotal records is to create a *clipboard or file folder with note cards*, one for each student, cascading down in an overlapping fashion so that as you are observing, simply flip to the child’s card that you are observing and jot down notes and, as you fill up cards, file them away into the student’s class file and place a new card in its place Hempeck (2009).

According to a design by Pierce and O’Malley (2009) anecdotal records can be produced by following three guidelines:-

- Describe a specific event, process, or product,
- Report rather than evaluate- save interpretation for later, and
- Relate the material to other known facts about the student.

Chen (2012) indicates that anecdotal records are useful. He was attracted to the scheme for it could help him to follow up the individual experience of any one student who would seem to merit personal observation as a result of his earlier interviews with him or her. The study further shows that this instrument has spaces for “other” categories which give it some flexibility and the user could easily change categories in the activity column, according to what he/she was looking for.

1.6 Portfolios

According to Stiggins et al. (2007), “the desire to capture and communicate the depth of student learning has been at the heart of portfolio use for years. A report card grade summarizes the story of achievement in one word but a portfolio gets at the full story, to help students, teachers, and others understand in depth one or more aspects of student learning. They are defined as files or binders which hold samples collection of individual student work (Hempeck, 2009, Papert, 2004). Nitko and Brookhart (2007) further add that for purposes of assessment, this collection is used either to present the student’s best work(s) or to demonstrate the students’ educational growth over a given time.

There are assessment portfolios and professional or instructional working portfolios (Robert & Pruitt, 2003). A professional portfolio is a thoughtful document demonstrating a teacher’s approach to teaching or an administrator’s approach to leadership. It offers a portrait of the educators practice over time and reflections about it (Martin-Kniep, 1999).Self-assessment and reflection are the most important functions of a portfolio in

terms of an educator's professional growth (Bullock & Hawk, 2001). Professional portfolios give candidates an edge in employment (Irby & Brown, 2000) as they may be used in lieu of the usual evaluation. Students benefit from the use of professional portfolios through improved instruction. Results with students should be an important part of the portfolio as the ultimate purpose of professional portfolio is to inform instruction and increase students' learning (Roberts & Priutt, 2003).

Portfolios provide a vehicle for preserving samples of outstanding teaching or leadership (Wolf, 1996). It helps teachers and administrators to collaborate to improve instruction and student outcomes (Van Wageningen & Hibbard, 1998). Collaborative portfolio development helps diminish isolation as educators share portfolios, discuss practice and exchange ideas about teaching and learning. The implications for future practice are perhaps the most important part of the reflection, for this is where learning is articulated (Irby & Brown, 2000).

Stiggins et al., (2007) feels that beyond their potential as a rich source of information, portfolios play a significant role in learning. Many times the responsibility for collecting, interpreting, and sharing portfolio contents falls to the teacher. However, we pass by a powerful learning experience when the *subject* of the portfolio (the student) is not also the *author*, for when students assemble and share portfolios, they perform acts of metacognition that deepen their ability to learn, their desire to learn, and the learning itself. Involving students in this work provides an ideal venue for getting them to take notice of, keep track of, and celebrate their learning. Collecting, organizing and reflecting on their own work build an understanding of themselves as learners and nurtures a sense of accomplishment. Taki and Heidari (2011) study indicate that portfolio-based writing assessment has a positive effect on language learning and writing ability. This study further, shows that it helps students' self-assessment and almost all students are satisfied with this method of assessment.

Hempeck (2009) advises that since it is neither a scrapbook nor a "damping ground" for all the student's accomplishments, items included therein are carefully and deliberately selected so that the collection as a whole accomplishes its purpose. On that note, Stiggins et al. (2007) caution that before we decide what goes into a portfolio, we must think about the story we want the portfolio to tell. The portfolio is only an organizing tool. To guide the selection of ingredients, they need a theme which could be struggle, achievement, competence, or celebration. What is the storyline—here is how this person completed each step of a project! See how this person has grown! Here is a picture of this person's level of achievement now! This person has attained these competencies! Here's is what this person is proud of! (Stiggins et al., 2007).

1.7 Checklists

A checklist, according to Nitko and Brookhart (2007), consists of a list of specific behaviors, characteristics or activities and a place for marking whether each is present or absent. Airasian and Russell (2008) however note some disadvantage -only two choices: criterion is performed or not, goal is met or not. There is no middle ground for scoring and no representation of extent. You may use a checklist for assessing a procedure a student uses, a product a student produces, or behaviors a student exhibits. Students may use checklists to evaluate their own performance. One can gather an array of information on a large group of students in an efficient manner. They are primarily useful if the behaviors are in a sequence or if all the subtasks that make up the complete performance can be listed.

1.8 Types of Checklists

Nitko and Brookhart (2007) give the following checklists.

1. *Procedure checklist*: A procedure checklist assesses whether the student follows the appropriate steps in a process or procedure e.g. the form represents both, the presence, or absence of each step and the sequence that a particular student used to perform a task.
2. *Product checklist*: A product checklist focuses on the quality of the thing a student makes. Products include drawings, constructed models, essays and term papers. These checklists identify the parts or other properties a product is supposed to have. You then inspect each product, checking whether those properties are present.
3. *Behavior checklist*: A behavior checklist consists of a list of discrete behaviors related to a specific area of a student's performance. For example, you may wish to identify the particular difficulties a student is having in the phonological, semantic, and syntactic aspects of spoken language. The behavior checklist might have items such as "uses only simple sentence structure" or "responds without delay to questions"
4. *Self - Evaluation checklist*: Students use a self-evaluation checklist to review and evaluate their own work. One could use the checklist students complete as a basis for a student teacher conference in which the teacher discusses a student's progress. A student must select six or seven completed tasks to put into the best works portfolio. A checklist can help a student evaluate each entry and decide what to put into the portfolio. The use of a checklist helps in overcoming lack of objective evaluation (Moore, 2009). It can also serve as a basis for discussing the entries with peers, parents or teachers. Because checklist focuses on portfolio entries, it focuses student's attention on the portfolio scoring rubric. However, the checklist is phrased in simpler and less formal language than the scoring rubric used by teachers.

1.9 Rating Scales

Rating scales are similar to checklists but differ in that they allow the observer to judge performance along a continuum rather than as a dichotomy (Airasian, 2000, 2001). Standards or criteria for evaluating a performance are created (Nitko & Brookhart, 2007). Each standard has levels of competence, and you rate students according to how well they complete the task i.e. level of degree of quality rather than as simply being present or absent. For example, to assess the quality of a student's oral presentation to the class one would probably identify several dimensions of a "good oral presentation" and then judge the degree to which a student demonstrates each of them. Good rating scales are carefully constructed measures that typically assess behaviors across relative frequency rating dimensions e.g. never, sometimes, always (Brown-Chidsey 2008).

Finally, having looked at how teachers should formatively assess and the various assessment methods, Bol et al. (1998) contended that although, there is an abundance of theory and opinion about why an alternative assessment should be used and how teachers might incorporate these assessments methods into their classrooms, there is relatively little research on the frequency with which teachers use various types of assessment methods and what they think about the different types of assessment.

2. METHODOLOGY

2.1 Research Design

This was a descriptive-comparative and evaluative study which tried to establish the use and usefulness of observational techniques for classroom-based ongoing assessment of students' performance in English in schools in Nandi Central District. For the students to perform well, it is necessary that English teachers and students periodically assess their progress. It is therefore important to ascertain whether the observational techniques are being put into good use. Improved performance and behavior during English lessons are criteria for judging the value of observational techniques. It is therefore, very important that teachers' and students' views on the use and usefulness of observational techniques in the teaching and learning of English be evaluated.

2.2 Population and Sampling Techniques

Secondary school teachers of English and students in Nandi Central District formed the target population for this study. According to the District Quality Assurance and standards Officer, there were 40 registered secondary schools in Nandi Central District where 26 were co-educational, 6 Girl's and 8 Boys' schools. These schools were also classified as Provincial 8, District 26.

To ensure proportionality, stratified sampling technique was used to select the sample. All schools in the District were categorized according to their type and listed in alphabetical order. These included Boys', Girls' and Mixed secondary schools. 30% of schools from each category were sampled to give a total of 12 schools (i.e. 12/40 - sample schools/total registered schools) as follows: Mixed Schools – eight (8), Boys Schools two (2) and Girls Schools- two (2).

Schools were randomly selected from each category to make the desired numbers. School names were written on 9 cm² paper for all schools in each category and then folded to hide their identity. The folded papers in order of 26, 8, and 6 were put on three open plates respectively as Mixed, Boys, and Girls School categories. 8, 2, and 2 schools were randomly selected from the plates. Teachers of English and the form three students in the selected schools automatically participated in the study. Cluster sampling was used to ensure all Form 3 students were involved. The sample consisted of 29 teachers teaching in the sampled schools and 1030 form three students. Form three students were used because they understand different teaching methods used and the various teaching and learning activities employed in English lessons. They have been exposed to the secondary school curriculum for two years hence, are more stable than the rest of the students. The teachers were TSC employees and trained BOG employees who had handled the form three classes.

2.3 Research Instruments

Both open and closed questionnaires were used whenever possible in order to maximize the number of questions which would be asked and for comprehensive feedback. This solicited for information ranging from general to specific areas related to the use of observational techniques to evaluate ongoing performance in English in secondary schools and their implications on the quality of teaching-learning process. The instrument was a two-part demographic and attitudinal questionnaire. The teachers' questionnaire contained demographic information about marital status, age of respondents, gender, qualifications and teaching experience, school where the respondent was teaching and home province. The structured items required respondents to select one response from the alternatives while the open type of the items consisted of statements that were aimed at revealing the attitudes and views of the teachers on particular issues. The students' questionnaire constituted structured questions on the observational techniques used by their teachers of English and their views on the use and usefulness of observational techniques to evaluate students' ongoing performance in English. Respondents were urged to respond to the questions in a classroom setup after which the questionnaires were collected.

2.4 Instrument Reliability

Reliability is the extent to which an item is able to trigger the same response or a respondent gives the same

answer consistently. To test reliability of the instrument, a pilot study was done in two secondary schools, one in Nandi South District and another in Uasin Gishu District to ensure that all items were reliable. Questionnaires were distributed accordingly to all the teachers of English and form three students. Cronbach's Alpha Coefficient was used to estimate the internal consistency of items and a coefficient above 0.6 was accepted as reliable. Analysis was done using Statistical Package for Social Science and a Cronbach's alpha coefficient of 0.948 was found for the student questionnaire. A reliability coefficient of 0.987 overall was found for teachers' questionnaire.

2.5 Data Gathering Procedures

Before getting in the data collection exercise, the National Council for science and technology issued the researchers with a research permit. An introduction letter was also given by the Nandi Central District Education Officer (DEO) that introduced the researchers to the sampled schools and facilitated collection of data in the secondary schools in the District. Data was collected in November 2011. The researchers first consulted the school principal, explained the purpose of the visit, and sought permission, after which was introduced to the teachers of English. The researcher talked to and, interviewed teachers, who then filled their questionnaires. The teacher accompanied the researchers to the form three classes. The questionnaire for the students was distributed and completed under classroom conditions with all the form three students involved. The researcher assured the respondents that their responses were for purposes of the research and would be treated with strict confidence.

2.6 Statistical Treatment of Data

The Mann-Whitney test was used to determine the difference in the perception of students and teachers on the teachers' use and usefulness of observational techniques to evaluate on-going performance in English because the responses were not normally distributed. The Kruskal-Wallis (K-W) test, a non-parametric statistical test, was used to determine the difference in the perception of teachers in mixed, boys, and girls schools on their perception on the use and usefulness of observational techniques because the responses were not normally distributed.

One way analysis of variance (ANOVA) was used to ascertain whether there was any difference among the perceptions of students on the use and usefulness of observational techniques in English instruction in boys only, girls only and mixed schools. Mann-Whitney test was used to show the difference in the perception of teachers while t-test was used to determine the difference in the perception of students on the use and usefulness of observational techniques in Provincial and District schools. The information collected from the field through questionnaires was analyzed using the Statistical Package for Social Sciences (SPSS).

3. RESULTS AND DISCUSSION

In this study we tested whether there was a significant difference in the evaluation of the use and usefulness of observational techniques in mixed, boys, and girls schools and between Provincial and District schools. The null hypothesis stated for this question is: *"There is no significant difference in the evaluation of the use and usefulness of observational techniques in the following schools(a) mixed, boys and girls schools and (b) Provincial and District schools as perceived by (i) the students and (ii) the teachers) at the level of significance of 0.05.*

If the null hypothesis is rejected at this level, it means that the probability is only 5% that the difference happened by chance. Statistically, this implies that the difference is real.

3.1 Comparison by School Composition for Teachers

The Kruskal-Wallis (K-W), a non parametric test, was used to determine the difference in the perception of teachers in mixed schools, boys' schools, and girls' schools on their perception on the use and usefulness of observational techniques.

Table 1 presents the test analysis of the difference in perception on rubrics, anecdotal records, portfolios, checklist rating scales and usefulness of observational techniques between teachers in Girls', Boys', and Mixed Schools. The K-W test yielded p-values of .377 (Rubrics), .342 (Anecdotal Records), .180 (Portfolios), .162 (Checklists), .886 (rating scales) and .947 (for usefulness of observational techniques)

Table 1: *Kruskal-Wallis Test: Test Analysis of Difference in Perception on Use of Rubrics, Anecdotal Records, Portfolios, Checklists, Rating Scales and Usefulness of Observation Techniques Between Teachers in Girls', Boys', and Mixed Schools.*

Observational Techniques	School Type	N	Mean	Std.	Chi-Square	p-value
Rubrics	Girls'	6	3.7222	1.12382	1.950	.377
	Boys'	8	3.2917	.72237		
	Mixed	15	3.0667	.86557		
Anecdotal Records	Girls'	6	3.1458	.84193	2.147	.342
	Boys'	8	3.0885	.46314		
	Mixed	15	3.5333	.67887		
Portfolios	Girls'	6	3.4167	1.04748	3.432	.180
	Boys'	7	2.4286	.95188		
	Mixed	15	3.0778	1.19335		
Checklists	Girls'	6	3.6389	.35616	3.637	.162
	Boys'	7	2.6190	1.01705		
	Mixed	15	3.3889	.99735		
Rating scales	Girls'	6	3.9259	.79401	.243	.886
	Boys'	7	3.6825	.60665		
	Mixed	15	3.5778	1.11760		
Usefulness of Observ Techq	Girls'	6	4.1818	.80289	.108	.947
	Boys'	8	4.2955	.26055		
	Mixed	15	4.0970	.77021		

which indicates that there is no statistically significant difference among the teachers in the three types of schools in their perceptions. The null hypothesis that there is no significant difference in the evaluation of the use and usefulness of observational techniques in girls' boys' and mixed secondary schools in the learning and evaluation of English is therefore accepted. Teachers felt that if the observational techniques are used, all students will benefit irrespective of whether they are learning in mixed, boys' or girls' schools. Madu and Kassanga (2005) in their research on Sex differences in the acquisition of English as a second language observed that the growing trend in establishing "unisex" or "mixed" schools is based on the principal that there should be no differentiation in the way girls and boys are taught. The underlying assumption is that girls and boys learn in the same way. Besides, different treatment of girls and boys might be perceived as discriminatory practice in an era of gender sensitivity. They further affirmed that the "myth" of female outshining males is not the full picture of beliefs on gender and language learning.

3.2 Comparison by School Composition for Students

3.2.1 Scoring Rubrics

Table 2 shows the analysis of comparing the mean of statements on use of rubrics between Girls, Boys, and Mixed schools. The analysis of variance results (ANOVA) shows that $F(2, 1013) = 27.809$; $p = .000$ which implies that there is a significant difference in the evaluation of the use and usefulness of rubrics between girls, boys, and mixed secondary schools in Nandi Central District as perceived by the students. Therefore, the null hypothesis is rejected. Girls' schools found rubrics more useful than their counterparts in boys' schools and mixed schools as indicated in their average mean ratings of 3.2749 2.6825 and 2.6955, respectively.

Table 2: *One-way ANOVA on Use of Rubrics*

School composition	Mean	df	Mean sq	F	Sig
Girls' school	3.2749				
Boys' school	2.6825				
Mixed school	2.6955				
Between Groups		2	33.831	27.809	0.000
Within Groups		1013	1.217		
Total	2.8471	1015			

$F(2, 1013) = 27.809$; $p = .000$

3.2.2 Anecdotal Records

Table 3 shows the analysis of the mean ratings of statements on anecdotal records.

Table 3: *One-way ANOVA on Use of Anecdotal Records*

School composition	Mean	df	Mean sq	F	Sig
Girls' school	3.3851				
Boys' school	2.2109				
Mixed school	3.0909				
Between Groups		2	94.176	112.490	0.000
Within Groups		1023	0.837		
Total	2.9702	1025			

$F(2, 1023) = 112.490; p = 0.000$

The above results show that $F(2, 1023) = 112.490; p = 0.000$. This finding reveals that the difference in the use of anecdotal records rating among girls' boys' and mixed secondary schools is statistically significant with a p – value of 0.000. This proves that there is a significant difference between the girls', boys' and mixed schools' evaluation of the use of anecdotal records. Therefore, the null hypothesis is rejected. Respondents from girls' schools had a higher mean rating of 3.3851 than the mean rating of those from boys' schools which is 2.2109 and from mixed schools which is 3.0909.

3.2.3 Portfolios

Table 4 illustrates test differences on the use of portfolios between girls', boys' and mixed schools.

Table 4: *One-way ANOVA on Use of Portfolios*

School composition	Mean	df	Mean sq	F	Sig
Girls' school	3.24				
Boys' school	1.82				
Mixed school	2.86				
Between Groups		2	134.63	133.34	0.000
Within Groups		1014	1.010		
Total	2.7258	1016			

$F(2, 1014) = 133.34; p = .000$

The p value of 0.000 indicates that there was a significant difference between use of portfolios in girls' boys' and mixed secondary schools as perceived by the students. Girls' schools had found portfolios more useful than their counterparts in boys' schools and mixed schools as indicated in their average mean ratings of 3.24, 1.82 and 2.86 respectively. This could have an impact on their performance and it is quite encouraging that girls already found them useful. Pearson (1999) noted that girls outperform boys in English. A closer look at their study habits reveal that Girls plan and organize more effectively than boys. Girls are also more likely to bring the right materials to the classroom. Girls seek out help more often than boys when they are struggling. When teachers comment on a students work, girls are more likely to be receptive to the suggestions than boys.

3.2.4 Checklists

Table 5 shows test of differences on use of checklists between girls', boys' and mixed schools.

P –value of .000 indicates that there was a significant difference between use of checklists and the evaluation of English in girls', boys' and mixed secondary schools in Nandi Central District. Students in girls, boys and mixed schools recorded means of 3.35 for girls' schools, 2.03 for boys' schools and 3.05 for mixed schools about their use of checklists. Boys' schools recorded the lowest means.

Table 5: *One-way ANOVA on Use of Checklists*

School composition	Mean	df	Mean sq	F	Sig
Girls' school	3.35				
Boys' school	2.03				
Mixed school	3.05				
Between Groups		2	119.942	127.949	0.000
Within Groups		1011	0.937		
Total	2.90	1014			

$F(2, 1014) = 127.949; p = .000$

Students in all the schools were in agreement that use of checklists has a significant influence on evaluation of on-going performance in English, recording an overall average mean of 2.90. Nitko and Brookhart, (2007) observes that students could use a self-evaluation checklist to review and evaluate their own work while the teacher could use the checklist students complete as a basis for a student teacher conference in which the teacher discusses student's progress in English.

3.2.5 Rating Scales

Table 6 shows test of differences on use of rating scales between girls' boys' and mixed secondary schools in Nandi Central District.

Table 6: *One-way ANOVA on Use of Rating Scales*

School composition	Mean	df	Mean sq	F	Sig
Girls' school	3.42				
Boys' school	2.31				
Mixed school	2.99				
Between Groups		2	76.779	81.777	0.000
Within Groups		1004	0.939		
Total	2.95	1006			

$$F(2, 1004) = 81.777; p = .000$$

The P-value of .000 indicates that there was a significant difference between use of rating scales and the evaluation of English in Girls', boys' and mixed secondary schools in Nandi Central District. Students in both girls, boys and mixed schools reported favorably about their use of rating scales recording average means of 3.42 for girls' schools, 2.31 for boys' schools and 2.99 for mixed schools. Students in girls', boys' and mixed schools were in agreement that use of rating scales has a significant influence in evaluation of on-going performance in English, recording an overall average mean of 2.95 though the girls had a high rating..

3.3 Usefulness of Observational Techniques

Table 7 shows the analysis of the mean ratings of statements on usefulness of observational techniques. The results shows that $F(2, 1008) = 9.092; p < .000$. The finding reveals that the difference in observational techniques rating among girls' boys' and mixed secondary schools is statistically significant at 0.05 with a p-value of .000. This reveals that there is a significant difference between the girls', boys', and mixed secondary schools evaluation of the usefulness of observational techniques. Therefore, the null hypothesis is rejected.

Table 7: *One-way ANOVA on Usefulness of Observational Techniques*

School composition	Mean	df	Mean sq	F	Sig
Girls' school	3.75				
Boys' school	3.53				
Mixed school	3.44				
Between Groups		2	8.781	9.092	0.000
Within Groups		1008	0.966		
Total	3.54	1010			

$$F(2, 1008) = 9.092; p = .000$$

The null hypothesis was rejected since the findings revealed that there was a significant difference in the use and usefulness of observational techniques between girls', boys' and mixed secondary schools in Nandi Central District as perceived by students. The students could have rejected the null hypothesis due to the general belief that girls outperform boys in English. Asch (1999) for instance noted that boys had not made so much progress in their areas of weakness-reading and writing, and that fewer boys than girls enrolled in almost every English course-Literature, composition and speech except on remedial English.

Madu and Kassanga (2005) agree that the "myth" of female outshining males is not the full picture of beliefs on gender and language learning. In reality, opinions are divided about which side of the gender divide outperforms the other in second language learning. Three main views have emerged overtime.

1. The first holds that females outperform males (Baker and MacIntyre, 2000; Schueller, 2000). The sociolinguistic finding that women use a higher frequency of the incoming forms in linguistic changes than men seems to have been used as a starting theory for Second language acquisition (SLA) research.
2. The second view states that males outperform females in various SLA activities and/or skills (Brantmeier, 2000; D'Ailly, 2002; Hassan, 2001, Kassanga, 1996).
3. The third view can be said to be a middle-ground position; authors in this group suggest that no

significant difference exists between the performance of males and females in SLA (e.g., Piske, Mackay & Fiege, 2001).

The differing views among researchers have to do with: whether boys and girls differ in the way they approach, learn, and perform in an L2: the nature and extent of gender differences, if any, in L2 learning and performance; the impact of these differences in instructed SLA; and practical of pedagogical measures to deal with them. Several studies (Gorman, et al., 1987; Gorman, et al., 1988) have reported that girls tend to outperform boys in reading and writing presumably because girls tend to have more positive feelings about reading and writing than boys. However, it seems that by the age of 15 years, girls lose confidence in their abilities as writers.

Some authors have, however, reported that boys performed better than girls in their studies (e.g., Hassan, 2001; Piske, Mackay & Fiege, 2001). Contrary to the findings, some authors found no significant difference of girls and that of boys (Jegade, 1994). For example, Jegede (1994) carried out a survey among high school students in Nigeria and reported that there was no significant gender difference in English Language academic achievement and motivation among participants; but that the student's English language performance could be reliably inferred from their level of achievement and motivation. The author therefore suggested that both boys and girls at the secondary school level were equally capable of mastering English.

In conclusion, the above review of studies on the performance of girls and boys illustrates the lack of consensus on the possible effect of gender on second language acquisition. It may be surmised that the culture context, tasks, and other research variables may explain the variety of findings. However, the results supports the growing trend in establishing "Unisex" or mixed schools based on the principal that there should be no differentiation in the way girls and boys are taught.

4. CONCLUSIONS AND RECOMMENDATIONS

Based on the above findings, the study concludes that teachers of English in Nandi Central District don't use observational techniques consistently and effectively, a reason for low performance in the subject. Teachers and students espoused the importance of observational techniques in equipping learning and teaching of English language to ensure quality performance in the subject. Scoring rubrics, anecdotal records and portfolios have similar impact when used in the evaluation of performance, while checklists, rating scales, and usefulness of observational techniques have different impact when used in the evaluation of English performance.

Students in the girls' schools found observational techniques more useful than the students in boys' and mixed schools, but the teachers were in agreement that the techniques are beneficial to all schools irrespective of their types. Teachers and students had varied opinions on the use and usefulness of observational techniques in Provincial and District schools. The students' accepted that only scoring rubrics could be used in both Provincial and District schools while the teachers agreed that all the observational techniques were useful in evaluating performance in both Provincial and District schools

References

- Airasian, P. W. (2000). *Assessment in the classroom: A concise approach (2nd ed.)*. Boston: McGraw-Hill.
- Airasian, P. W. & Russell, M. K., (2008) Recording Classroom observations, Retrieved from <http://www.slideshare.net/Dollyo/recording-techniques-2008> The new Xperia™ X8 Android TMSmartphon.
- Baines, L. (2008) *A teachers' guide to multisensory learning: Improving literacy by engaging the senses*. Alexandria, Virginia USA: Association for Supervision and Curriculum Development.
- Baker, S. C., & MacIntyre, P. D., (2000) The role of gender and immersion in communication and second language orientations. *Language learning*, 50(2):311-340
- Bol, L., Stephenson, P. L., O'Connell, A. A., Nunnery, J. A., (1998) Influence of experience, grade level and subject area on teachers' assessment practices. *Journal of Educational Research*, Vol. 91, Retrieved from <http://www.questia.com/googleScholar.qst?docId=97820106>
- Boyd-Batstone, P., (2004) *Focused anecdotal records assessment; a tool for standard-based, authentic assessment: teachers can use this technique to develop common ground for authentic assessment in a standards-based environment*. Retrieved July 10, 2009, from <http://www.accessmylibrary.com/article-1C1-125148236/focused-anecdotal-records-assess>
- Brantmeier, C. A. (2000). The relationship between readers' gender, passage contents, comprehension and strategy use in reading Spanish as a second language. *Dissertation Abstracts International section A: Humanities and social sciences* 61(4-A):1376.
- Brown-Chidsey, R. (2008). *Assessment for intervention: A problem solving approach*. New York: The Guilford Press.
- Bullock, A. A., & Hawk, P. P., (2001) *Developing a teaching portfolio: A guide for pre-service and practicing teacher*. Upper Saddle River, NJ: Prentice Hall.
- Chen, L., (2012) Aspects on learner-based classroom observational techniques.

- Journal of English language teaching*, Vol. 5 (5). Retrieved from <http://www.dx.doi.org/10.5339/elt.v5n5p148>.
- D'Ailly, H (2002) *A distinctive cultural and gender differences in children's interests and effort in learning: The impact of choice and testing paper presented at the annual meeting of the American Educational research Association*, new Orleans LA, April 1-5, 2002.
- Danielson, C (2002) *Enhancing student achievement: A Framework for school improvement*. Alexandria, VA: ASCD.
- Gorman, T. P., White, J. and Brooks, G. Maclure, M. and Kispad, A. (1988). *Language performance in school: Review of APU Language monitoring 1979-1983*. London: HMSO Department of science/ Department of Education for Northern Ireland/ Welsh office.
- Gronlund, N.E. & Linn, R.L., (1990) *Measurement and evaluation in teaching* (6thEd). New York: Macmillan.
- Hassan, B. A., (2001) *Extraversion/Introversion and gender in relation to the English pronunciation accuracy of Arctic speaking college students*. Research Report (33pp).
- Hempeck, N., (2009) *Methods for documenting student progress: Strategies for Assessing and Keeping Records in the Classroom*: Retrieved from <http://www.Suite101.Com/profile.cfm>
- Jegade, J. O. (1994). Influence of motivation and gender on secondary school students' academic performance in Nigeria. *Journal of social psychology*, 134(5):695-698
- Kenya National Examination Council (2007).
- Meyer, H.W., (2009) "When you just get a mark and a nasty comment, what's that Called?", "Summative assessment". Creating an enabling environment for formative assessment. *Southern African Linguistics and Applied Language Studies*. 27 (2): 215-228.
- Madu, B. N. & Kasanga, L, A (2005) Sex differences in the acquisition of English as a second Language. Retrieved from <http://www.ajol.info/index.php/gab/article>
- Irby, B. J., & Brown, G. (2000) *The career advancement portfolios*. Thousand Oaks, CA. 'Corwin.
- Kasanga, L. A. (1996) Effect of gender on the rate of interaction. Some implications for second language acquisition and classroom practice. *I.T.L Review of applied linguistics*, 111/112:155 – 192.
- Kellaghan, T. & Greany, V (2001) *Using assessment to improve the quality of education*. Paris: UNESCO.
- Kutner, M. Greenberg, E. Jin, Y. Boyle, B. Hsli, Y. & Dunsleavy, E, (2007) *Literacy in everyday life*. Washington D.C.: US. Department of Education
- Martin-kniep, G. O., (1999) *Capturing the wisdom of practice*. Alexandria, VA: Association for supervision and curriculum development.
- Moni, K., Kraayenoord, C.V., & Baker, C. (1999) English teachers' perceptions of literacy assessment in the first year of secondary school. *Australian Journal of language and literacy*, Vol. 22. Retrieved from <http://www.questia.com>
- Moore, K. D., (2009) *Effective instructional strategies: From theory to practice* (2ndEd.). USA: SAGE Publications, Inc.
- National Assessment of Adult Literacy., (2005) *A first look at America's adults in the 21st century*. Washington, DC: US Department of Education.
- National Centre for Education Statistics. (2005) *A profile of the American high school sophomore in 2002*. Washington, DC: US Department of Education.
- Nitko, A. J., & Brookhart, S. M., (2007) *Educational assessment of students* (5th Ed.). Ohio: Pearson Merrill Prentice Hall.
- Papert, S., (2004) *Evaluation: Building Blocks of Constructivism* (2004) Retrieved from <http://www.saskschools.ca>.
- Petty, G., (2004): *Teaching today: A practical guide* (3rd Ed.) United Kingdom: Nelson Thomas Ltd.
- Pierce, L. V. & O'Malley J. M., (2009) *Performance and portfolio assessment for language minority students*. Retrieved from <http://www.finchpark.Com/Courses>
- Piske, T., Mackay, I. R. A. and Flege, J. E. (2001) Factors affecting degree of foreign accent in an L2: A review *Journal of phonetics*. 29(2):191-215.
- Roberts, S. M., & Pruitt, E. Z., (2003) *Schools as professional learning communities collaborated activities and strategies for professional development*. California: Corwin press, Inc.
- Schueller, J.M. (2000) The effects of two types of strategic training on foreign language reading comprehension: An analysis by gender and proficiency. *Dissertation Abstracts International section A: Humanities and social sciences*, 60(7-A):2472.
- Simmons, L.K. (2009) *Secondary English assessment*. Retrieved from <http://en.wikibooks.org/wiki>
- Stiggins, R. J., Arter, J., Chapuis, J., & Chapuis, S., (2007) *Classroom assessment for student learning, doing it right -using it well*. New Jersey: Pearson Prentice Hall.
- Taki, S., & Heidari, M., (2011) The effect of using portfolio-based writing assessment

on language learning: The case of young Iranian EFL learners. *Journal of English language teaching*, Vol. 4 (2). Retrieved from <http://journaldatabase.org>.

Van-Wagenen, L., & Hibbard, K. M., (1998) Building teacher portfolios. *Educational leadership* 55 (5) 26-29.

Watson, T. S., & Steege, M. W., (2003) *Conducting classroom-based functional behavioral assessment: A practitioner's guide*. New York: The Guilford Press.

Yi, Q., (2012) Empirical study of formative evaluation in adult ESL teaching. *Journal of English language teaching*, Vol. 5 (2). Retrieved from <http://journaldatabase.org/articles>