

Test and Item Response Theories and School Environment as Assessment Practice Factors among Science and Mathematics Teachers in Secondary Schools in Eastern Uganda

¹Deborah Manyiraho,^{*1,2} Dennis Zami Atibuni, PhD and ¹David Kani Olema, PhD
¹Busitema University, Uganda

²University of Johannesburg, Auckland Park Kingsway, South Africa

*Correspondence Author: dmanyiraho@yahoo.com

Abstract

This study sought to establish the effect of knowledge of the Classical Test Theory (CTT) and Item Response Theory (IRT) and school assessment environment on assessment practice among teachers of science and mathematics subjects in Eastern Uganda Secondary Schools. The study assessed the levels of knowledge and application of CTT and IRT in assessment, examined the suitability of school environment for assessment and established the influence of school environment and knowledge of CTT and IRT on teachers' engagement in assessment. A census of 307 teachers of science and mathematics subjects attending SESEMAT training in Eastern Uganda participated in the study. The results revealed that the teachers were engaged in assessment ($M = 17.04$, $SD = 2.00$) and had moderate levels of knowledge of CTT ($M = 10.19$, $SD = 2.23$) and IRT ($M = 17.5$, $SD = 3.50$). Their levels of application of CTT ($M = 28.08$, $SD = 3.85$) and IRT ($M = 6.86$, $SD = 1.47$) were also moderate. The teachers reported that their schools had somewhat conducive environments for assessment ($M = 14.37$, $SD = 3.44$). In addition, school environment affected teachers assessment practices most ($\beta = .211$, $t = 7.212$, $p < .05$), knowledge of CTT also influenced teachers assessment practice, but less than the influence by environment ($\beta = .112$, $t = 4.969$, $p < .05$). In conclusion, enhancing the levels of knowledge and application of CTT and IRT as well as improving school assessment environment are paramount for meaningful engagement in assessment by teachers. The study recommended pre-service and in-service training of the teachers in CTT and IRT in addition to schools improving environments for effective teacher engagement and quality assessment.

Keywords: Assessment, Item Response Theory, Classical Test Theory, Assessment Environment, engagement

Introduction

The teaching and learning process is incomplete without assessment. Assessment, the methods or tools that teachers and students use to evaluate, measure, and report the academic readiness, learning progress and skills acquisition, plays a key role in realizing educational needs of students (Jabbarifar, 2009). Assessment involves gathering, analyzing and interpreting information about students and their academic progress using tests, quizzes, observation checklist, oral questions and examinations (National Research Council, 2001). Assessment can be diagnostic, formative or summative. To realize the central role of assessment in the process of teaching and learning, a meticulous

consideration of assessment practices of teachers is mandatory in an education system that is accountable to society. However, these assessments are often done informally based on student feedback and simple measures such as average scores (Sudol & Studer, 2010). The informal evaluation of assessments may in one way or another affect the quality of assessment. Consequently, teachers need to have adequate knowledge as well as skills to construct classroom based or teacher made tests to evaluate learning outcomes and achievement.

However, many teacher-made tests are not built according to international guidelines and best practices, such as application of theories, and are

never evaluated for performance to check whether they are producing reliable results (Thompson, 2016). This may be as a result of poor school environment for assessment, low teacher engagement in assessment and inadequacies in knowledge of test theories that teachers receive during training and career development. Such inadequacies consequently limit their ability to set and score test items, analyze test results and give appropriate feedback to stakeholders (National Research Council, 2010).

Therefore, this study focused on the assessment of the levels of knowledge and application of Classical Test Theory (CTT) and Item Response Theory (IRT) in assessment by teachers of science and mathematics subjects. Since assessment is carried out in a school environment, the study further examined the suitability of school environments for assessment and established the influence of school environment and knowledge of CTT and IRT on teachers' engagement in assessment.

School Environment and Teachers' Assessment

Teachers' engagement in assessment, especially in terms of science teachers interacting with materials and resources that enhance their skills to assess learners, needs to take place in a motivating environment. Such resources include text books, online resources, apparatus, and colleagues, which are usually in the school setting where the teachers serve. The availability and accessibility to these resources for teachers constitute an enabling environment for professional interaction. In addition, an enabling school environment for assessment is a facilitator for teachers to demonstrate their knowledge on CTT and IRT. Some school environments are characterized by inadequacy of resources/materials for assessment and understaffing in science departments, which implies that teachers cannot consult each other when developing assessments. This, to some extent, may negatively affect the quality of assessment.

In attempt to mitigate such inadequacies, the Ministry of Education and Sports (MOES) in Uganda instituted mandatory Secondary Science and Mathematics Teachers' (SESEMAT) trainings which have served as an avenue of increasing science teachers' interaction. This has enabled teachers of science and mathematics subjects to network nationally, share best classroom practices and improve teachers' engagement in assessment (SESEMAT Regional Report, 2019). The SESEMAT

regional report explains that in Tororo SESEMAT Region, the trainings have further improved teachers' skills in assessment through school based SESEMAT Activity Regional Based (SARB) initiatives where science teachers are encouraged to analyze tests and exams that they set for their students and conduct remedial lessons in topics that students have performed poorly.

According to An and Yung (2014), improved classroom assessment and evaluation, under the active management of teachers, serves important professional development purposes since the information resulting from such evaluations provide teachers with valuable feedback about their instructional effectiveness. Science teachers can therefore use classroom assessment to develop and improve their professional knowledge and skills which includes knowledge and practice of CTT and IRT which are commonly used in classroom assessment.

Classical Test Theory

Schuwirth and Vander Vleuten (2011) regard Classical Test Theory (CTT) as a theory of assessment that is premised on the central assumption that the observed score is a combination of the true score and an error score ($O = T + e$). CTT is the most widely used theory in learner assessment since it is applicable to most topics (Conejo, Guzmán, & Trella, 2016). When applying CTT, one is expected to always set a parallel test for the students and calculate the test-retest correlation to ensure that the test is reliable. It is also necessary that the teacher makes an analysis of the items during assessment. According to Thompson (2016), CTT provides some methods for evaluating items based on simple statistics like proportions, correlations and averages which teachers commonly apply. Therefore, for quality assessment, teachers need to be equipped with knowledge about the CTT so that they can apply it whenever need arises. For instance, CTT is helpful in straight forward assessment situations such as the standard open-ended or multiple-choice test. Teachers of science therefore employ CTT in construction of open ended and multiple-choice questions that form the compulsory parts of examinations (Uganda National Examinations Board [Uganda National Examination Board, 2018]).

However, CTT poses a difficulty in assessing individuals' abilities by using test items. The theory does not take into consideration the different

abilities of the learners during item writing. Secondly, the classical test theory bases its reliability on parallel tests which are difficult to achieve in reality, especially where science teachers are few and/ or are faced with big class sizes and heavy teaching loads as is often the case in Tororo SESEMAT Region. Additionally, missing values in CTT are difficult to handle during both test development and subject scoring, which makes it difficult to analyze items that have random missing data.

Item Response Theory

Item response theory (IRT), on the other hand, is concerned with accurate test scoring and development of test items (An & Yung, 2014). An and Yung note that one advantage of the item response theory is that, the item characteristics and the personal abilities formulated by distinctive parameters can be derived for a population, enabling the scores from that population to be compared directly (invariant property of IRT). According to Schuwirth and Vander Vleuten (2011), IRT is important in overcoming the problem of separating test difficulty effects from candidate group effects by estimating item difficulty independent of student ability, and student ability independent of item difficulty.

Furthermore, measurement precision can be depicted by information curves in IRT which can be treated as a function of the latent factor conditional on the item parameters. They can be calculated for an individual item (item information curve) or for the whole test (test information curve). Missing values in IRT are easy to handle during both test development and subject scoring, making it straightforward to analyze items that have random missing data. Because of this, IRT is widely used in education to calibrate and evaluate items in tests, questionnaires and other instruments, and to score subjects on their abilities, attitudes or other latent traits (An & Yung, 2014). According to Thompson (2016), item response theory is much more powerful than the classical test theory, but only works with sample sizes numbering in the hundreds or larger. If science and mathematics teachers have knowledge of this theory, then they can apply item response theory in large scale testing while for classroom sized samples or other small-scale situations ($N < 100$), classical test theory can be applied. As such, IRT is relevant to teachers during assessment of their subjects especially in large testing bodies or large schools. However, it proves

less useful for classroom testing as the students are often few.

Wrenn and Wrenn (2009) note that for effective transition from theory to practice, teachers require adequate knowledge and skill. This holds true even for educational assessment. However, the application of test theories during educational assessment seems to receive little emphasis in professional development programs, yet assessment is a core aspect of teachers' routine. According to SESEMAT training reports, a lot of emphasis is put on classroom pedagogy and subject content (SESEMAT Regional Report, 2018) at the expense of valid and useful assessment practices administered in a conducive environment on the basis of concrete theories. Generally, there is a dearth of investigation into teachers' assessment practices including incorporation of classical test and item response theories in test construction which could affect students' grades. This means that efforts to improve the pedagogical practices of science and mathematics teachers as emphasized in SESEMAT workshops will continue to be based on decisions that are not informed by empirical evidence. Thus, the prevailing poor grades in science at Uganda Certificate of Education (Akwei, 2017) in the Eastern region of Uganda are likely to persist, exacerbating the national divide in academic performance. This study provides information to stakeholders in education on how school environment and knowledge of CTT and IRT can be harnessed to improve the quality of assessment. The study may also inform teachers on the importance of basing assessment practice on theory so that assessment is more useful in terms of reliability, validity and effectiveness. The study was guided by the following objectives:

1. To assess the relative levels of knowledge and application of CTT and IRT in assessment by teachers of science and mathematics subjects in Eastern Uganda.
2. To examine the suitability of school environment for assessment in science and mathematics subjects in Eastern Uganda.
3. To establish the influence of school environment and knowledge of CTT and IRT on teachers' engagement in assessment.

Research Methodology

The study adopted a convergent parallel mixed methods design employing both quantitative and qualitative approaches to provide a greater depth

and therefore understanding of the topic and analysis of data from the study population at a specific point at the same time (Creswell, 2014). Participants in the study were a census of 307 teachers of science and mathematics attending mandatory SESEMAT in-service trainings in eastern Uganda. A census of the participants was taken to obtain robust data that would be more reliable.

Data for the study were collected using a self-report questionnaire with three sections. Section A contained background information on the school type and the participant. Section B contained questions related to school assessment environment, teachers' engagement in assessment and level of knowledge and application of CTT and IRT. Section C contained open ended questions to seek explanation on how the teachers carried out assessment and their knowledge of CTT and IRT. The items for measuring suitability of school assessment environment, teachers' engagement in assessment and levels of knowledge and application of CTT and IRT were self-constructed and subjected to a pilot test among teachers of arts subjects in Tororo. The psychometric properties were then determined. The reliability of the scale measuring knowledge and application of CTT and IRT was found to be .70 meaning that the scale was consistent. The content validity index (CVI) was .88, meaning the instrument was relevant. The items were scored on a 5-point Likert scale; 1 (*Strongly Disagree*) to 5 (*Strongly Agree*). The scores of items in the same scale were summed up to generate the overall scale score. The minimum scale score was obtained from multiplying the number of items by one, and maximum obtained by multiplying the number of items by five. Generally, higher scores on each scale indicated higher levels of knowledge and skills in either theory.

Quantitative data were analyzed using SPSS (version 20) generating both descriptive and inferential statistics (*i.e.* Mean standard deviation, percentages, frequencies and correlation coefficients). To achieve objectives one and two, mean and standard deviation of the overall variable score were determined. The results were interpreted as low, moderate and high depending on the score ranges obtained after dividing the range of the overall scale score into three equal parts. Achievement of objective three involved first determining teachers' level of engagement in assessment followed by establishing the correlation between the variables as interpreted by Evans

(1996), and finally running a multiple linear regression model. The qualitative responses from the open ended items in the questionnaire were reported using interpretative phenomenological analysis approach in the format "verbatim quote" (Participant number, date). The verbatim quotes generally corroborated the quantitative findings.

During the planning, collection and processing of data, informed consent of the Participants was sought before administering questionnaires. It was made known to the Participants that their participation was voluntary and they were free to withdraw from the study at any time or may not answer questions they were uncomfortable with. Due respect was given to the Participants' privacy and confidentiality, in which case their names and identities were concealed was granted. Objectivity was one of the guiding principles to the study to avoid bias as well as display a high level of confidentiality with data collected from the Participants. The academic documents such as journal articles, reports, books, and book chapters used in the study as sources of information were all duly credited by citing in text and referencing.

Analysis and Discussion

The analysis of data begins with the presentation of demographic characteristics of respondents and then moves into descriptive analysis as well as inferential statistics in which a hypothesis was tested to determine the interrelationships among the variables under investigation.

Demographic Profiles of Respondents

Demographic information obtained shows that majority of the participants were from schools in the rural (61.6%), which were mainly day (57.7%), government-aided (74.9%), mixed (87%), USE (61.2%) schools, as detailed in Table 1.

The type of school usually determines the assessment environment in terms of availability of materials and resources. Besides, the schools had learners from diverse categories which poses challenges to the teacher during setting. For instance, one teacher said, "It is hard to make choice of questions that can fit all the categories of learners" (Participant183, Jan 2019). Class sizes were averagely between 50 and 100(62.5%) students as reported by the participants. This explains why most of the teachers gave big class sizes as a challenge to carrying out quality assessment in their subjects. According to one

Participant, “big number of learners in class gives hard time to mark, analyze and give feedback in the

shortest time possible” (Participant 171, Jan 2019).

Table 1: Background Information of the Participants (N = 307)

Characteristic	Category	n	%
Location of School	Rural	189	61.6
	Urban	118	38.4
School Type	Day	177	57.7
	Boarding	72	23.5
	Both day and boarding	58	18.9
School Ownership	Government-aided	230	74.9
	Private	77	25.1
Policy of Funding	USE	188	61.2
	Non USE	119	38.8
Sex of Students	Mixed	267	87.0
	Single sex	40	13.0
Average Class Size	<50	41	13.4
	50 – 100	192	62.5
	100 – 150	51	16.6
Gender of Participant	Male	273	88.9
	Female	34	11.1
Highest Level of Qualification	A Level	6	2.0
	Diploma	126	41.0
	Degree	161	52.4
Age of Participant	Post graduate	14	4.6
	20 – 29	134	43.6
	30 – 39	94	30.6
	40 – 49	53	17.3
Major subject of Specialization	>49	26	8.5
	Biology	95	30.9
	Chemistry	39	12.7
	Mathematics	96	31.3
Title of Participant	Physics	77	25.1
	Teacher	209	68.1
	Head of Department	86	28.0
	Director of Studies	9	2.9
Teaching Experience	Deputy Head teacher	3	1.0
	< 10	183	59.6
	10 -20	91	29.6
	20 – 30	24	7.8
	> 30	9	2.9

In terms of gender, the participants were mainly male (88.9%) implying that very few women teach sciences in eastern Uganda. According to Dee (2006), the gender of a teacher determines the effectiveness of the teacher in terms of treating boys and girls differently in the classroom. Such differential treatment results in differences in overall academic performance. Dee further notes that on a national assessment, boys score higher in math and science, while girls score higher in reading. Majzub and Rais (2010), in their Malaysian study, observed that boys fared badly in the public examinations compared to girls. On further investigation, they discovered that boys were doing

less well in schools compared to girls, and had lower chances to qualify to enter the universities than girls. Dee (2006) also notes that students are more engaged, behave more appropriately and perform at a higher level when taught by a teacher who shares their gender. In short, girls have better educational outcomes when taught by women and boys are better off when taught by men. This implies a possibility that classroom assessments by teachers in eastern Uganda will mainly favor boys.

However, the teachers were qualified mostly at diploma and degree levels (93.4%) implying that most of them were technical in pedagogy and assessment practices. In addition to this, majority of

the participants were below 40 years (74.2%) and were mainly classroom teachers (68.1%) with teaching experience of less than 10 years (59.6%). This means that they were actively engaged in classroom work and assessment of students. Due to their age, their cognitive abilities could maintain functional independence such as learning new skills (Clark, Freedberg, Hazeltine, & Voss, 2015). The subjects of specialization for the participants were nearly evenly distributed across the subjects except for chemistry which had the lowest percentage (12.7%).

Knowledge and Application of CTT and IRT by Teachers of Science and Mathematics

The first objective of the study sought to establish teachers' level of knowledge and practice (application) of CTT and IRT which were computed and the results are summarized in Table 2. Quantitative findings presented in Table 2 show that science and mathematics teachers in Eastern Uganda had moderate levels of knowledge of CTT and IRT and they moderately applied these theories during assessment.

Table 2: Descriptive Statistics for Teachers' Knowledge and Application of CTT and IRT

	Minimum	Maximum	<i>M</i>	<i>SD</i>	Level
Level of knowledge of CTT	3.00	15.00	10.19	2.23	Moderate
Level of Practice of CTT	5.00	25.00	17.49	3.51	Moderate
Level of Knowledge of IRT	8.00	40.00	28.08	3.85	Moderate
Level of Practice of IRT	2.00	10.00	6.86	1.47	Moderate
Overall Level of Test Theories	20.00	86.00	62.61	7.80	Moderate

However, the qualitative findings indicate that majority of the teachers did not have concrete knowledge of the test theories. When asked which theory they use while setting items, Participant145 (Jan 2019) said, "theory of bloom's taxonomy which includes analysis, assessment, knowledge and comprehension." Some teachers did not have true knowledge of the test theories. They confused CTT and IRT with theories of their subject content. For instance, one Participant said, "In setting test items, I employ a theory that relates, for instance, a set of three square numbers" (Participant150, Jan 2019). Another Participant said, "I employ explanation of the basic concepts of particular processes in organisms" (Participant41, Jan 2019). The third category of Participants acknowledged honestly that they did not employ any test theories. For instance, Participant 167 said that, "during item setting I choose questions commonly set in past papers" (Participant167, Jan 2019). As advocated by Schuwirth and Vander Vleuten (2011), teachers need to have a sound knowledge of theories underlying assessment in order to develop good quality assessment. Otherwise, there is a danger of ignoring the possibilities, limitations and underlying goals and assumptions of the assessment which could result in over- or underestimations of the reproducibility of the assessment. Participants 137 and 168 (Jan 2019) suggest that refresher trainings and workshops aimed at teachers' assessment practices could be of help in overcoming shortfalls of knowledge in assessment. Besides, the

assessment theories investigated in this study provide a rich and well-organized knowledge base for effective practice not only in assessment per se but also in effective teaching and learning using interactive strategies.

Together with Baird and Black (2013), the authors believe that teachers need to have a high level of knowledge of the theories of assessment in order to construct tests that can discriminate well amongst the test-takers. In this case, not only would they be "consumers of already composed test items" (Participant244, Jan 2019), but they would critically and creatively come up with novel contextualized items tailored to the appropriate levels of difficulty and discrimination to foster learning in their own settings.

To Examine the Suitability of School Environments for Assessment

The second objective of the study sought to examine the suitability of school environments for assessment. The teachers indicated that school environments were moderately suitable with the mean scores ranging from 4.00 to 20.00 ($M = 14.37$, $SD = 3.44$). This means that the majority of teachers agreed that their school environments were suitable for assessment, implying that they had enough resources and personnel for quality assessment among others. However, responses from the open ended questions indicated that big class sizes affected the quality of assessment. For instance, one teacher said, "due to the high number of learners, I

may not be able to analyze each learner's performance and give feedback, so I usually give a general one" (Participant 42, Jan 2019). Inadequacies in school assessment environment is also implied in suggestions proposed by some Participants on how to improve their assessment practices. For instance, some said, "by equipping laboratories, buying reagents and increasing manpower" (Participants 137, 140, Jan 2019).

However, as noted by Simmonds (2017), where the environment is not supportive for teachers' effective undertaking of their duties, they need to have the courage, knowledge and commitment to depart from traditional discourses and expectations and engage in complicated conversations towards social transformation. In this way, they would be positive change agents for the community. As asserted by Islahi and Nasreen (2013), the effectiveness of education is very much dependent on the effectiveness of its teachers.

A disabling environment does not promote the acquisition of the requisite learning outcomes of most values that enable human flourishing - as citizens, as workers, as family and community members and as fulfilled individuals -who are

enabled to continue learning in the rapidly changing, information- and technology-rich environment (James, 2006). Therefore, teachers should not only adapt to the environment they find themselves posted in, but they should strive to adapt the environment to bring about the desired transformation in their learners through developing approaches to teaching and assessment that encourage critical thinking and interactive problem-solving. Helping teachers to become more effective may therefore mean retooling them attitudes of harnessing their environments to enhance learning.

I Knowledge of CTT and IRT, Assessment Environment and Teachers' Engagement

Objective three sought to establish how teachers' knowledge of the test theories and school assessment environment influenced science teachers' level of engagement in assessment. The teachers were highly engaged in assessment with scores ranging from 7.00 to 20.00 (M = 17.04, SD = 2.00). The teachers invested quality time in assessment, assisted students who faced challenges in their classes and were interested in participating in continuous professional development programs aimed at improving their assessment practices.

Table 3: Correlation between Assessment Environment, Knowledge of CTT and IRT and Teachers' Engagement

Variables		School assessment environment	Knowledge of CTT	Knowledge of IRT	Teachers' engagement in assessment
School assessment environment	Pearson Correlation	1	.131*	.115*	.405**
	Sig. (2-tailed)		.022	.045	.000
	N	307	307	307	307
Knowledge of CTT	Pearson Correlation	.131*	1	.392**	.339**
	Sig. (2-tailed)	.022		.000	.000
	N	307	307	307	307
Knowledge of IRT	Pearson Correlation	.115*	.392**	1	.202**
	Sig. (2-tailed)	.045	.000		.000
	N	307	307	307	307
Teachers' engagement in assessment	Pearson Correlation	.405**	.339**	.202**	1
	Sig. (2-tailed)	.000	.000	.000	
	N	307	307	307	307

Note. ***Correlation is significant at the 0.001 (2 tailed), **Correlation is significant at the 0.01 (2 tailed), *Correlation is significant at the 0.05 (2 tailed)

This high level of engagement in assessment probably was as a result of most teachers being professionally qualified for their job. Islahi and Nasreen (2013) argue that teaching is not restricted to lecturing and that teachers are expected to motivate, inspire, explain, engage, understand and guide the students for their all-round development. Teachers are further expected to attract students

towards content, concepts, attitudes, values, knowledge and skills under the set frame of pre-determined goals. This requires of them a certain critical level of professional proficiency, educational efficiency and social sufficiency. Exhibiting these values culminates in their overall level of professional engagement which is tagged to the socially valued objectives of education in the setting

in which they work, and the environmental determinants that enhance or inhibit their performance.

To establish whether teachers' knowledge of test theories and school assessment environment influenced their engagement in assessment, Pearson Product Correlation Coefficients were computed and a multiple linear regression model was run. The following criteria were used to interpret nature of existing relationships: $\geq .70$ = strong relationship; $\geq .50$ = moderate relationship and $\leq .50$ = weak relationship.

Table 3 shows that there is a positive yet weak relationship between school assessment

environment and teachers' engagement (.405), between knowledge of CTT and teachers' engagement (.339) and between knowledge of IRT and teachers' engagement (.202).

The generally weak correlation coefficients suggest that apart from school assessment environment, knowledge of CCT and knowledge of IRT, there are other factors that influence teachers' level of engagement in assessment. As noted by Wrenn and Wrenn (2009), there is a high likelihood that teachers face a difficulty in making transition from theory to practice.

Table 4: Assessment Environment, Knowledge of CTT and IRT and Teachers' Engagement

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.499 ^a	.249	.242	1.74177

a. Predictors: (Constant), Knowledge of IRT, Knowledge of CTT, School assessment environment

Table 5: ANOVA on Assessment Environment, Knowledge of CTT and IRT and Teachers' Engagement

Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	305.381	3	101.794	33.554	.000 ^b
	Residual	919.225	303	3.034		
	Total	1224.606	306			

a. Dependent Variable: Teachers' engagement in assessment

b. Predictors: (Constant), Knowledge of IRT, Knowledge of CTT, School assessment environment

Table 6: Coefficients of Independent and Dependent Factors

Model		Unstandardized Coefficients		Standardized Coefficients Beta	t	Sig.
		B	Std. Error			
1	(Constant)	10.064	.882		11.406	.000
	School assessment environment	.211	.029	.363	7.212	.000
	Knowledge of CTT	.112	.023	.270	4.969	.000
	Knowledge of IRT	.024	.024	.054	1.001	.318

a. Dependent Variable: Teachers' engagement in assessment

Furthermore, a multiple linear regression model was performed to ascertain whether school environment and knowledge of CTT and IRT were the factors that predicted teachers' level of engagement in assessment. The results are presented in Tables 3 and 4.

Results in Table 4 indicate an R value of 0.499 showing that there is a moderate positive relationship between the stated predictors and science teachers' engagement in assessment. The results further indicate that 24.9% of the changes in teachers' engagement in assessment are influenced by changes in school assessment environment and knowledge of CTT and IRT altogether. The Sig. F

change value ($F < .001$) as shown in Table 4 indicates that the model was correctly specified and that the effect of the predictors on teachers' engagement is significant at a 1% significance level.

From Table 6, it can be seen that every unit improvement in school assessment environment results in a 21.1% increase in teacher engagement ($B = .211, t = 7.212, p < .05$). Similarly, a unit increase in the levels of knowledge CTT causes 11.2% improvement in teachers' engagement in assessment ($B = .112, t = 4.969, p < .05$). However, a unit increase in knowledge of IRT does not produce a significant change in teacher engagement in assessment ($B = .024, t = 1.001, p$ greater than .05).

These results imply that teachers will strive to put into practice the CTT knowledge rather than IRT knowledge they learn from training colleges and in-service trainings if the environment for assessment is conducive. Both school assessment environment and CTT knowledge contribute up to 32.3% effect in teacher engagement in assessment, meaning that other factors prevail that also influence the teachers' practices in assessment. As noted by Wrenn and Wrenn (2009), for effective transition from theory to practice, teachers require training in self-awareness, knowledge acquisition and skill building. Further, the teachers need to have skills in relationship building, exploring or probing, empowering, challenging and gaining and utilizing knowledge from their personal experience and practice as teachers.

As observed by Islahi and Nasreen (2013), the effectiveness of secondary school teachers is determined by several factors including their gender that influence their job performance. These authors note that teachers need to have a complex set of skills, insight, intelligence, knowledge, management, competence, dynamism and diligence to meet the challenges of the classroom in order to effectively apply the theories they gained from training. This set of skills is demanded more so in effective assessment for learning, which is central and integral to teaching and learning.

James (2006) asserts that teachers' assessment practice is inevitably influenced by external assessment processes and teachers often use these assessments as models for their own, even if they do not use them directly. James further argues that by using models of assessment borrowed from elsewhere, teachers submit to the temptation of subscribing, uncritically or unwittingly, to the theories of learning on which those assessments are based. This study indicates that teachers do have moderate levels of knowledge and practice of CTT and IRT to underpin their assessment practice, and they are able to articulate them as shown by the high level of engagement reported in the study. As noted by James, there ceases to be a boundary between the different theories when it comes to borrowing of tenets of the theories in developing assessments for learners.

Conclusions and Recommendations

This part presents the conclusions of the study and then gives the recommendations.

Conclusions of the Study

In conclusion, teachers of science and mathematics in Eastern Uganda do engaged in assessment with moderate levels of knowledge and application of CTT and IRT. The school assessment environments were moderately suitable for assessment. Moreover, school assessment environment and knowledge of CTT influenced teachers' engagement in assessment up to 32.3%. The positive correlation between the levels of knowledge of CTT and IRT and teachers engagement in assessment is evidence that the independent variables affect teachers' engagement in assessment procedures. However, the results bring to the fore the fact that there are other environmental variables that may affect the teachers' engagement while applying the assessment theories.

Recommendations of the Study

The study recommends that for quality feedback on the effectiveness of instruction, the teachers need to be equipped with adequate knowledge on assessment beyond moderate. They should also be enabled to apply this knowledge in their classroom situations. This can be achieved through targeting their pre-service and regular SESEMAT in-service trainings towards such ends. The pre-service teacher training curriculums need to emphasize the theory and practice of assessment and evaluation as a course. The study also recommends that schools should improve their environments for assessment by providing teachers with adequate materials for teaching, learning and assessments in addition to ensuring that class sizes are manageable. Finally, while the study showed that there are factors besides knowledge of assessment theories and suitability of assessment environment which also affect assessment practice; these factors need to be investigated in future studies.

Reference

- Akwei, I. (2017). *Poor performance in science: Ugandan education minister blames teachers.* Africa News. www.africanews.com
- An, X., & Yung, Y. (2014). Item response theory: What it is and how you can use the IRT procedure to apply it. *Paper SAS 364-2014*. <https://support.sas.com/resources/papers/proceedings14/SAS364-2014.pdf>
- Baird, J. A., & Black, P. (2013). Test theories, educational priorities and reliability of public examinations in England. *Research*

- Papers in Education*, 28(1), 5-21. <https://doi.org/10.1080/02671522.2012.754224>
- Clark, R., Freedberg, M., Hazeltine, E., & Voss, M. W. (2015). *Are there age-related differences in the ability to learn configural responses?* *PLoS One* <https://doi:10.1371/journal.pone.0137260>
- Conejo, R., Guzmán, E., & Trella, M. (2016). The SIETTE automatic assessment environment. *International Journal of Artificial Intelligence in Education*, 26(1), 270-292.
- Creswell, J. W. (2014). *Research design: Qualitative, Quantitative and mixed methods approach*, 4th ed. Lincoln. Sage Publications, Inc.
- Dee, T. S. (2006). The why chromosome: How a teacher's gender affects boys and girls. *Education Next*, 6(4), 68-76. <https://cepa.stanford.edu/content/why-chromosome-how-teachers-gender-affects-boys-and-girls>.
- Evans, J. D. (1996). *Straight forward statistics for the behavioural sciences*. Pacific Grove, CA. Brooks/Cole Publishing.
- Islahi, F., & Nasreen, N. (2013). Who Make Effective Teachers, Men or Women? An Indian Perspective. *Universal Journal of Educational Research*, 1(4), 285-293. <https://doi.org/10.13189/ujer.2013.010402>
- Jabbarifar, T. (2009). The importance of classroom assessment and evaluation in educational system. *Proceedings of the 2nd International Conference of Teaching and Learning (ICTL, 2009)*. Malaysia.
- James, M. (2006). Assessment, teaching and theories of learning. In J. Gardner (Ed), *Assessment and Learning* (1st ed., pp. 47-60). London: Sage. <https://doi.org/10.13140/2.1.5090.8960>
- Majzub, R. M., & Rais, M. M. (2010). Boys' underachievement: male versus female teachers. *Procedia-Social and Behavioral Sciences*, 7, 685-690. <https://doi.org/10.1016/j.sbspro.2010.10.093>
- National Research Council (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC. The National Academics Press. <https://doi.org/10.17226/10019>.
- National Research Council (2010). *State Assessment Systems: Exploring Best Practices and Innovations: Summary of Two Workshops*. Washington, DC: The National Academies Press. <https://doi.org/10.17226/13013>.
- Schuwirth, L. W. & Vander vleuten, C. P. M. (2011). General overview of the theories used in assessment. *AMEE Guide in Medical Education*, 57. *Medical Teacher*, 33 (10). p. 783 – 797. <http://doi.org/10.3109/0142159X.2011.611022>
- SESEMAT Regional Report (2018). *Report by regional trainers on the January 2018 INSET at Tororo Girls' School*. Unpublished.
- Simmonds, S. (2017). Teachers as curriculum leaders: towards promoting gender equity as a democratic ideal. *Educational Research for Social Change*, 6(2), 16-28. <https://doi.org/10.17159/2221-4070/2017/v6i2a2>
- Sudol, L. A., & Studer, C. (2010). Analyzing test items: Using item response theory to validate assessments. In proceedings of the 41st ACM Technical Symposium on Computer Science Education, *SIGCSE'10*, 436 – 440. New York, NY, USA, 2010. ACM.
- Thompson, N. A. (2016). Introduction to classical test theory with CITAS. *Assessment Systems for Good Measure*. https://assess.com/docs/Thompson_2016_Classical_Test_Theory_with_CITAS.pdf
- Uganda National Examinations Board (2018). *Report on candidates work UCE 2017*. UNEB. Kampala
- Wrenn, J., & Wrenn, B. (2009). Enhancing learning by integrating theory and practice. *International Journal of Teaching and learning in higher education*, 21(2), 258-265. <http://www.isetl.org/ijtlhe/>.