

## **EFFECT OF COGNITIVE DRILL THERAPY ON THE DEVELOPMENT OF MATHEMATICS TEACHERS' TEST ITEM WRITING SKILLS**

John J. AGAH<sup>1</sup>, Lawrence I. AGUELE<sup>2</sup>

<sup>1</sup>University of Nigeria, Faculty of Education (Nigeria)

<sup>2</sup>Ambrose Alli University Ekpoma, Faculty of Education (Nigeria)

john.agah@unn.edu.ng, profaguele@gmail.com

### **Abstract**

Test item writing is one real-life problem that poses a great challenge to teachers and often impacts negatively on students' achievement as well as decisions taken about the students. It is pertinent to use cognitive drill therapy (CDT) to resolve test item writing challenges experienced by teachers. The authors adopted a Non-equivalent control group research design. A total of 83 teachers participated in the study and constituted the study sample. Mathematics Test item Writing Skills Rating Scale (MTIWSRS) was the instrument used for data collection. The internal consistency of MTIWSRS was determined using Cronbach Alpha and a reliability coefficient of 0.83 was obtained. Data was analyzed using analysis of covariance. It was found that cognitive drill therapy (CDT) had a more significant effect on participants' test item writing scores. The researchers concluded that cognitive drill therapy (CDT) enhanced the Mathematics teachers' item writing skills in Enugu State, Nigeria. CDT was a good professional development programme for training Mathematics teachers in Enugu State, Nigeria.

**Keywords:** Cognitive drill, Item writing skills, Experience, Gender

### **1. Introduction**

The inability of students to attain the desirable level of achievement in internal and external examinations globally and in Nigeria, in particular, has been attributed to the students, teachers, schools, curricula factors and the assessment approaches employed. The assessment process is a source of worry for the researchers in this article. The teachers who teach and examine students in the classroom appear to have some challenges in writing good items, as such they ignore task that measures higher-order thinking skills during item writing. This could be one of the reasons why most students cannot sufficiently display higher-order skills in areas like creativity, abstract thinking, and critical reasoning, (Aydın, Sarier, & Uysal, 2012). The challenges that teachers face in writing items to measure higher-order thinking skills can also be considered among the reasons for students' failure as well as their inability to create and discover new things (İnceçam, Demir, & Demir, 2018; Barak, & Dori, 2009).

Increasing students' mathematical achievement in internal and external examinations is considered to be closely related to the cognitive level of teacher questions used in teaching processes or formative assessment (Çalık & Aksu, 2018; Sahin & Kulm, 2008) and the quality of teacher-designed tests and measurement instruments used in examinations (Aygün, Baran-Bulut, & İpek, 2016). For example, Çalık and Aksu (2018) state that teachers and prospective teachers prefer to ask questions at lower thinking levels during teaching. The writing of good items to measure pupils/students' learning in all subjects especially Mathematics has posed a serious challenge to teachers. Most teachers see it as a difficult task and they simply draw items or questions directly from past examinations, textbooks or even notebooks when assessing learners.

This act helps some learners who have access to the questions to pass well in internal examinations but fail at external examinations, hence the need for further professional development of Mathematics Teachers in item writing skills. The training will enhance Mathematics teachers' development and in turn, improve their skills in assessment.

Assessment is a major element that enhances students' thinking and develops their skills (Alkhateeb, 2019). It is reported that classroom teachers in most countries of the world and Nigeria in particular lack adequate knowledge and skills to write good items for assessment and it is a global phenomenon (Çakan, 2004; Şata, 2016). Teachers equally have some challenges on how to use assessment to measure higher-order thinking competencies and they are also unable to use assessment sufficiently to support learning. This may indicate the need for more quality learning opportunities on assessment. However, Professional Development activities on assessment are limited, and participation is rather low in Turkey (Aydın, Selvitopu, & Kaya, 2018; Kitchen et al., 2019) and even Nigeria.

Teacher development (TD) entails a comprehensive, sustained, and intensive approach to improving teachers' effectiveness in raising learners' achievement and it is regarded as Professional development. It is activities that aim to develop an individual's skills, knowledge, expertise, and other characteristics as a teacher (OECD, 2009). Teacher development has positive effects on students' learning outcomes and achievement (Loyalka, Popova, Li, & Shi, 2018). It is a form of in-service training that could improve the teachers' skill set and knowledge base. Participation in seminars, conferences, online sharing, belonging to professional groups, and other methods are some of the ways mathematics teachers can support their professional growth. This assertion corroborates Sharma's (2015) position that in formal settings, professional development programmes are often disseminated through seminars, conferences, workshops, and sometimes occur in informal conditions, through interactions across online platforms, and peer group discussions among others. The rate of active professional development (PD) activities such as observation visits, peer and/or self-observation and coaching is low globally (OECD, 2019). The fact that teachers had been trained to teach thinking skills to some extent, of which they were less often trained to assess (Stiggins, Griswold, & Wikelund, 1989) makes it pertinent to involve them in professional development.

Some Mathematics teachers have difficulty in determining higher or lower-level thinking skills and creating test items for higher-order thinking (Thomson, 2008). As such, teachers need to be trained with the right programs (Driana & Ernawati, 2019). Teachers of mathematics work extremely hard to make the subject engaging and interesting for their students. To improve the quality of educational delivery, teachers must occasionally participate in professional development activities. Majority of the teachers lacked competencies for preparing quality classroom tests, particularly on the use of Table of Specification as well as lacked professional support on how to prepare matching items, short answers, and multiple-choice test items. Effective teacher development has a wide range of benefits, including improved teaching methods, higher levels of student engagement, and better learning outcomes.

Mathematics teachers need to inculcate, sustain, and have a strong understanding of item writing skills, especially higher-order thinking and make it an integral part of the classroom (Sebastian, 2020). Teachers' understanding of Bloom's Taxonomy which divides learning objectives into three domains: cognitive, affective, and psychomotor is essential. The cognitive domain includes six hierarchical categories: knowledge, comprehension, application, analysis, synthesis, and evaluation. Knowledge is at the lowest cognitive level, while evaluation is at the highest level. Learning objectives and questions can be phrased using the specific verbs that are linked with each category of the cognitive domain.

Mathematics Teachers' knowledge of the six levels of the cognitive domain is essential to higher-order thinking skills (HOTS) and it fosters students' interest in mathematics as well as creates continuous learning that will instil creativity among individuals (Abdul et al, 2017).

Moreover, teachers must pay equal attention to questioning skills that provide constructive feedback or ask constructive follow-up questions (Sebastian, 2020 & Nicol, 1999). It has been observed that most teachers are not good at constructing tests in their various subject areas. As such, most teachers hastily duplicate questions from any previous exam. Majority of test questions created by teachers are poorly designed and lack psychometric features. This revealed why Quansah et al. (2019) said that teachers have limited skills in the construction of end-of-term examinations and suggested organizing workshops for teachers regularly to sharpen their skills in effective test construction practices.

In constructing a good test, Nworgu (2015) outlined some stages which include content analysis, review of instructional objectives, development of test blueprint/table of specification, and item writing among others. As pointed out earlier, item writing practices pose a great challenge to classroom teachers. The practice involved writing the stem of the item using verbs to present problems, stating questions without irrelevant information, keeping the language simple and clear, avoiding tricky items, writing the question without giving too many clues to answers and so on. Teachers who do not possess sound skills in item writing are bound to produce questions that increase test anxiety among students. This will in turn impact students' academic performance (Ali & Abdul-Wahab, 2022; Ibrahim, 2018). Researchers have also argued that test construction among teachers has not been encouraging (Ali & Abdul-Wahab, 2022; Hamafyelto et al., 2015; Kazuko, 2018). Furthermore, the test construction skills of teachers are inadequate, especially those with few years of teaching experience (Ebinye, 2011) and worse for non-professional teachers (Ololube, 2008). However, Adodo (2014) found out that years of experience do not make any significant difference in teachers' knowledge of test construction procedures.

Well-written test items contribute to accurate assessment of students' mastery and in turn, improve the precision of judgment about students' learning and achievement. Anderson (2017) found that experienced teachers were more adept at aligning test items with curriculum objectives, ensuring that assessments accurately measured what students were supposed to learn. Some experienced teachers write test items that measure a wide range of student responses and misconceptions (Smith, 2019). They are more likely to design items that probe for specific misconceptions, resulting in assessments that provide valuable insights into student understanding (Johnson, 2020). Williams, (2021) opines that their item writing practices align with evolving pedagogical approaches and assessment methods that permit more innovative and application-oriented item designs. Moreso, they incorporate/engage in reflective practice, analyzing the outcomes of their assessments and iteratively improving their item writing skills in regular self-assessment and feedback processes, leading to the refinement of their item writing skills over time (Martinez, 2018). The teachers play a mentoring role, guiding novice educators in test item writing skills development, and giving professional mentorship skills to foster the growth of novice educators' item writing abilities (Turner, 2019). Although experienced teachers have a deeper understanding of curriculum standards and learning objectives as well as possess the qualities enumerated by researchers, most of them still display unsatisfactory skills in item writing.

There are various gender issues related to teaching and learning at schools at all levels (Paudel, 2020). Gender refers to a network of beliefs, personalities, traits values, behavior and activities differentiating women and men through a process of social construction that has several distinctive features (OECD, 2018, UNDP, 2006). It is a situation in which stereotypical roles or attitudes are attributed to males or females (Halpern, Straight, & Stephenson, 2011).

It noted that gender gaps do not fluctuate between items or sections of the exam (Reardon & Ho, 2015), and female teachers were less likely to perceive themselves as proficient in test item writing skills compared to their male counterparts (Brown, 2017). This means that gender inequalities exist across test item writing skills. Some subtle biases could unintentionally manifest in the wording and framing of questions. This can impact the perceived difficulty and accessibility of items for different genders (Smith, 2020).

Researchers have found that there exists a significant difference between males and females in the knowledge of test construction procedures whereas others report no differences. In the reports, some are in favour of males for tests with more multiple-choice items and more in favour of females for tests with more constructed-response items (Reardon, Kalogrides, Fahle, Podolsky, & Zárate, 2018). Furthermore, gender differences may be noticed in irrelevant skills required by different item types (Scheiber, Reynolds, Hajovsky, & Kaufman, 2015). Also, the test item format creates gender differences in achievement across state tests that weigh various dimensions of learning using overall scores (Reardon, Kalogrides, Fahle, Podolsky, & Zárate, 2018; Taylor & Lee, 2012). It was found that cognitive behavioral active engagement training was efficacious in the improvement of test items construction skills among primary school teachers in Nigeria (Ede, et al, 2021). It is essential to determine how gender could moderate the item writing skills of teachers when trained with Cognitive Drill Therapy and lecture methods. The essence is to ascertain whether the disparity associated with performance due to experience and gender could be addressed with Cognitive Drill Therapy.

A lot of teachers feel uncomfortable when they are told to submit quality test items for examination in their schools. Some of them feel uneasy when developing worthy exam questions that a superior person is to vet before it is administered to students. This creates anxiety which is characterized by constant fear and worry in various school settings. Teachers in this category are found to be deficient or incompetent in writing quality test items in Mathematics and require specific training through cognitive drill therapy. Cognitive drill therapy (CDT) is a new approach to the treatment of stimulus-bound anxiety (Kumar et al. 2012). The therapy is based on the theories of conditioning, cognitive appraisal, and linguistics. CDT has given promising results in many research studies (Verma, Arya, Kandhari & Kumar, 2018; Arya, Verma & Kumar, 2017),

CDT conceptualizes fear as having four components: O- Objects of fear; B- Body Mind Reactions; S- Safety Behaviors and D- Danger Perceptions (Arya, Verma, Kumar, & Mishra, 2018). The object of fear (carrying out content analysis and preparing test blueprint), causes physiological and psychological responses (trembling, nervousness, irritability, lack of concentration among others), teachers cope by avoiding the situation (copying questions from textbook/notebooks, past questions) as the teacher feels pressurized due to various reasons (test validators will make fun of him/her, weak skills in test item writing, etc.). It is therefore pertinent to find out the effect of Cognitive Drill Therapy on the development of Mathematics Teachers' test item writing Skills.

### **Research questions**

1. What is the mean rating of Mathematics Teachers' test item writing skills scores when exposed to Cognitive Drill Therapy and lecture method?
2. What is the influence of experience in the mean rating of Mathematics Teachers' test item writing Skills scores?
3. What is the influence of gender in the mean rating of Mathematics Teachers' test item writing Skills scores?

### Hypotheses

1. There is no significant difference in the mean rating of Mathematics Teachers' test item writing Skills scores when exposed to Cognitive Drill Therapy and lecture method.
2. There is no significant influence of teachers' experience in the mean rating of Mathematics Teachers' test item writing Skills scores.
3. There is no significant influence of teachers' gender in the mean rating of Mathematics Teachers' test item writing Skills scores.

### 2. Methods

The quasi-experimental design was used for the study, specifically, the non-equivalent control group design was employed. The area of the study was Enugu State, Nigeria. The sample size was 83 Mathematics teachers consisting of 41 male and 42 female teachers using a purposive sampling technique. The instrument that was used to collect data for the study was Mathematics Test item Writing Skills Rating Scale (MTIWSRS). MTIWSRS was a 15-item five-point rating scale of Excellent = 5, very good = 4, good = 3, poor = 2 and very poor = 1. It was face-validated by two experts from Measurement and Evaluation and one in the Mathematics Education unit of the Science Education Department, Faculty of Education, University of Nigeria Nsukka (UNN). The teachers were observed and rated by seven (7) experts using MTIWSRS. The internal consistency of MTIWSRS was determined using Cronbach Alpha and a reliability coefficient of 0.83 was obtained. In this study, cognitive drill therapy was done in face-to-face interaction. The teachers in the CDT group were first asked to relax by taking deep breaths. When they felt relaxed and calm, participants were asked to imagine themselves in the challenged situation in Mathematics. At this point, the drill of what to do (item writing skills) was presented to them. The teachers began the sessions, by imagining a particular situation and repeating the drill statement severally, then a pause of 20-30 seconds or larger was given before resuming. The therapy continued from the application of the drill statement until the fear of writing good items diminished. This session of CDT lasted from 45 minutes to more. In the present study, CDT intervention was carried out in two centres. Teachers were also given CDT homework to repeat drill statements twice a day, and they were encouraged to use cognitive drill therapy when they felt uncomfortable or stressed. The training lasted for eight weeks. Before the commencement of the treatment, MTIWSRS was administered as a pretest and after the treatment, it was again administered as a posttest. The data collected were analyzed using SPSS, and the research question was answered with mean and standard deviation. The hypotheses were tested at 0.05 level of significance using Analysis of Covariance (ANCOVA). The pre-test scores were used as covariates for the post-test scores.

### 3. Results

The result presented in Table 1 shows the pretest and posttest mean rating of Mathematics Teachers' test item writing skills scores for Cognitive Drill Therapy and lecture method. The result showed that the CDT group had a pretest mean score of ( $\bar{x} = 25.86$ ,  $SD = 4.17$ ) and a posttest mean score of ( $\bar{x} = 48.34$ ,  $SD = 3.12$ ). The mean difference was 22.48. The result also showed that the LM group had a pretest mean score of ( $\bar{x} = 25.97$ ,  $SD = 5.41$ ) and a posttest mean score of ( $\bar{x} = 31.38$ ,  $SD = 3.64$ ) with a mean difference of 5.67. The result reveals that the CDT group had a higher mean gain in Mathematics item writing skills than their LM counterparts.

Table 1. Mean and Standard deviation of ACT and CT groups

		Pretest		Posttest		
Treatment	N	$\bar{x}$	SD	$\bar{x}$	SD	Mean Difference
CDT Group	44	25.86	4.17	48.34	3.12	22.48
LM Group	39	25.97	5.41	31.38	3.64	5.67

CDT = Cognitive Drill Therapy; LM = Lecture Method

The result presented in Table 2 shows the pretest and posttest mean rating of Mathematics Teachers' test item writing skills scores for 5-years & below and 6-years & above. The result showed that the 5-year & below group had a pretest mean score of ( $\bar{x} = 25.05$ ,  $SD = 4.13$ ) and a posttest mean score of ( $\bar{x} = 40.06$ ,  $SD = 8.34$ ). The mean difference was 15.10. The result also showed that the 6-years & above group had a pretest mean score of ( $\bar{x} = 26.85$ ,  $SD = 5.25$ ) and a posttest mean score of ( $\bar{x} = 40.70$ ,  $SD = 10.04$ ) with a mean difference of 13.85. The result reveals that the 5-years & below group had a slightly higher mean gain in Mathematics item writing skills than their 6-years & above counterparts.

Table 2. Mean and Standard deviation of 5-years &amp; below and 6-years &amp; above

		Pretest		Posttest		
Experience	N	$\bar{x}$	SD	$\bar{x}$	SD	Mean Difference
5-years & below	43	25.05	4.13	40.06	8.34	15.10
6-years & above	40	26.85	5.25	40.70	10.04	13.85

The result presented in Table 3 shows the mean rating of males and females in Mathematics Teachers' test item writing skills. The result showed that the male teachers had a pretest mean score of ( $\bar{x} = 25.92$ ,  $SD = 5.48$ ) and a posttest mean score of ( $\bar{x} = 39.12$ ,  $SD = 9.04$ ). The mean difference was 13.20. The result also showed that the female group had a pretest mean of ( $\bar{x} = 25.90$ ,  $SD = 4.00$ ) and a posttest mean score of ( $\bar{x} = 41.59$ ,  $SD = 9.19$ ) with a mean difference of 15.59. The result reveals that the female group had a slightly higher mean gain in Mathematics item writing skills than their male counterparts.

Table 3. Mean and Standard deviation of male and female groups

		Pretest		Posttest		
Gender	N	$\bar{x}$	SD	$\bar{x}$	SD	Mean Difference
Male	41	25.92	5.48	39.12	9.04	13.20
Female	42	25.90	4.00	41.59	9.19	15.59

The result in Table 4, shows that the main effect due to treatment was significant ( $F = 504.23$ ,  $\rho = 0.00$  and  $\eta_p^2 = 0.872$ ), the main effect due to experience was not significant ( $F = 1.10$ ,  $\rho = 0.298$  and  $\eta_p^2 = 0.02$ ), the main effect due to gender was not significant ( $F = 0.97$ ,  $\rho = 0.756$  and  $\eta_p^2 = 0.001$ ), the interaction effect for treatments and experience was not significant ( $F = 3.930$ ,  $\rho = 0.051$  and  $\eta_p^2 = 0.05$ ) and the interaction effect for treatments and gender was equally not significant ( $F = 1.396$ ,  $\rho = 0.241$  and  $\eta_p^2 = 0.019$ ). This implies that there is a significant difference in the mean rating of Mathematics Teachers' test item writing Skills scores when exposed to Cognitive Drill Therapy and lecture method in favor of the CDT group.

The partial eta square,  $\eta_p^2$  (effect size) of 0.875 means that 87.2 % of the increase in the mean test item writing Skills values of Mathematics Teachers was due to the effect of training they received.

Table 4. ANCOVA result on item writing Skills for CDT and LM groups

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Dec
Corrected Model	6023.460 <sup>a</sup>	8	752.932	66.175	.000	.877	
Intercept	3773.659	1	3773.659	331.66	.000	.818	
PreIWS	.147	1	.147	.013	.910	.000	
Groups	5737.084	1	5737.084	504.23	.000	.872	S
Experience	12.512	1	12.512	1.100	.298	.015	NS
Gender	1.108	1	1.108	.097	.756	.001	NS
Groups * Experience	44.718	1	44.718	3.930	.051	.050	NS
Groups * Gender	15.889	1	15.889	1.396	.241	.019	NS
Experience * Gender	3.015	1	3.015	.265	.608	.004	NS
Groups * Experience * Gender	5.146	1	5.146	.452	.503	.006	NS
Error	841.962	74	11.378				
Total	142157.000	83					
Corrected Total	6865.422	82					

a. R Squared = .877 (Adjusted R Squared = .864)

#### 4. Discussion

It was found that CDT and LM significantly enhanced the test item writing skills of Mathematics teachers in favour of the CDT group. The findings revealed that 87.2 % of the increase in the mean test item writing values of secondary school teachers was due to the effect of the training they received. Consequently, teachers of the secondary schools who participated in the training have gained skills on how to write good items in mathematics. The teacher went through a well-organized training which was targeted at developing their skills in item writing and the outcome was encouraging. These results collaborated with OECD (2009) position on improving teachers' effectiveness in raising learners' achievement by involving them in activities that aim to develop their individual skills, knowledge, expertise, and other characteristics. Developing teachers' skills in item writing has positive effects on students' learning outcomes and achievement (Loyalka, Popova, Li, & Shi, 2018). The findings in this research complement previous research that focused on test item writing or test construction. The finding is in consonant with Ede, et al, (2021) who reported that cognitive behavioral active engagement training was efficacious in the improvement of test items construction skills among primary school teachers in Nigeria, and giving professional mentorship skills fostered the growth of novice educators' item writing abilities (Turner, 2019). Thus, the use of CDT was potent in helping Mathematics teachers to write the stem of the item using verbs to present problems, write questions without irrelevant information, avoid writing questions that will test reading skills more than content knowledge in mathematics, and arrange items in order of difficulty among other skills.

Other findings of the study were the removal of differences due to experiences and gender. It was found that there was no significant gender-related difference among teachers after being exposed to CDT. The main effect due to experience was not significant and the interaction effect for treatments and gender was equally not significant. Although variations were observed in mean scores based on experience and gender, the values were not substantial for significant differences. This means that the therapy was able to remove the gender gap as well as experience differences observed by other researchers. This result may have been influenced by the fact that both male and female teachers took the training seriously. Also, the novice teachers were conscious of their involvement in training with more experienced teachers who may have been mentoring their professional growth and as such, were at their best. This finding negates the report that female teachers were less likely to perceive themselves as proficient in test item writing skills compared to their male counterparts (Brown, 2017). The finding was in accord with the report that years of experience do not make any significant difference in teachers' knowledge of test construction procedures (Adodo, 2014) and also not in tandem with the assertion that test construction skills of teachers are inadequate, especially those with few years of teaching experience (Ebinye, 2011) and worse for non-professional teachers (Ololube, 2008). The inadequacies of teachers in test construction that were earlier observed by researchers may be due to the use of inappropriate training programmes employed in developing them.

Overall, the result is indicative that Mathematics teachers who participated in the study have acquired new knowledge, skills, and value; can display change in behaviors in writing quality test items and use what they have learnt to impact positively on the schools resulting in a better outcome. They can be hired by individuals, schools, states, commissions, as well as examination bodies to participate in quality test item writing whenever the need arises. Through this research, an aspect of human development in Nigeria has been resolved.

### Conclusion

Cognitive drill therapy as well as the lecture method enhanced Mathematics teachers' item writing skills in Enugu state. CDT and LM were good training programmes for Mathematics teachers, but CDT was a better and more effective way of developing Mathematics teachers' item writing skills.

**Recommendations/Future directions:** Cognitive drill therapy (CDT) is potent in enhancing Mathematics teachers' item writing skills and is appropriate for teachers' training. It was therefore recommended that Cognitive drill therapy (CDT) should be used for teachers' professional development. Subsequent studies should focus on the use of CDT to reduce teachers' anxiety associated with item writing and teachers of various subjects be involved. Focusing on teachers of various subjects as participants for future studies will provide researchers with more insight into the potency of CDT in enhancing the item writing skills of teachers.

### References

- Abdullah, A. H., Mokhtar, M., Noor, D. A., Ali, D. F., Tahir, L. M., Umar, H. & Kohar, A. (2017). Mathematics teachers' level of knowledge and practice on the implementation of higher-order thinking skills (HOTS). *EURASIA Journal of Mathematics Science and Technology Education* 13(1), 3-17 doi:10.12973/eurasia.2017.00601a
- Adodo, S. S. (2014). An evaluation of secondary school teachers' competency in evaluating students' cognitive and psycho-motor achievement in basic science and technology. *Journal of Emerging Trends in Education Research and Policy Studies (JETERAPS)*, 5(3), 48-53.



- Ali, A. T. & Abdul-Wahab, I. (2022). Enhancing test development and item writing skills of subject and vocational teachers in government-owned junior secondary schools in Jigawa State, Nigeria. *Journal of Teaching and Teacher Education* 11(1), 1-12. <http://dx.doi.org/10.12785/jtte/110101>
- Alkhateeb, M. A. (2019). Assessing eighth-grade mathematics teachers and textbook in embodying thinking levels. *International Journal of Instruction*, 12(1), 371-388.
- Anderson, L. (2017). Influence of teachers' years of experience on test item alignment with curriculum. *Journal of Mathematics Education*, 32(2), 145-162.
- Arya, B., Verma, S., Kumar, R., & Mishra, R. (2018). Role of cognitive drill therapy in treatment of evaluation anxiety: A case study. *International Journal of Scientific and Research Publications*, 8(5), 297-300. <http://dx.doi.org/10.29322/IJSRP.8.5.2018.p7743>
- Aydın, A., Sarier, Y., & Uysal, Ş. (2012). The comparative assessment of the results of mathematical literacy in terms of socio-economic and socio-cultural variables. *Education and Science*, 37(164), 20-30.
- Aygün, B., Baran-Bulut, D., & İpek, A. S. (2016). The analysis of the primary school mathematics exam questions according to the math taxonomy. *Turkish Journal of Computer and Mathematics Education*, 7(1), 62-88.
- Barak, M. & Dori, Y. J. (2009). Enhancing higher order thinking skills among in-service science teachers via embedded assessment. *Journal of Science Teacher Education*, 20(5), 459-474. doi: 10.1007/s10972-009-9141-z
- Brown, A. (2017). Gender differences in mathematics teachers' test item writing skills. *Gender and Education*, 25(4), 523-540.
- Çakan, M. (2004). Comparison of elementary and secondary school teachers in terms of their assessment practices and perceptions toward their qualification levels. *Ankara University Journal of Faculty of Educational Sciences*, 37(2), 99-114. doi:10.1501/Egifak\_0000000101
- Çalık, B. & Aksu, M. (2018). A systematic review of teachers' questioning in Turkey between 2000-2018. *Elementary Education Online*, 17(3), 1548-1565. doi:10.17051/ilkonline.2018.466389
- Driana, E. & Ernawati, E. (2019). Teachers' understanding and practices in assessing higher order thinking skills at primary schools. *Journal of Teaching & Education*, 1(2), 110-118. Retrieved from: <https://journals.umkt.ac.id/index.php/acitya/article/view/233>
- Ebinye, P. O. (2001). Problems of testing under the continuous assessment programme. *Journal of Quality Education*, 4(1), 12-19.
- Ede, M. O., Agah, J. J., Okeke, C. I., Chuks, Z. O., Oguguo, B. C. E., Agu, P. U., ... & Manafa, I. F. (2021). Effect of cognitive behavioral active engagement training on test item construction skills among primary school teachers in Nigeria: implication for educational policymakers. *Medicine*; 100:36(e26876).
- Hamafyelto, R. S., Hamman-Tukur, A., & Hamafyelto, S. S. (2015). Assessing teachers' competence in test construction and content validity of teacher-made examination questions in Commerce in Borno State, Nigeria. *Education*, 5(5), 123-128. doi: 10.5923/j.edu.20150505.01
- Ibrahim, A. (2018). Improving assessment and evaluation skills of public-school teachers in Jigawa State, Nigeria. *Journal of Education and Practice*, 9(11), 22-32. [www.iiste.org](http://www.iiste.org)

- Inçeçam, B., Demir, E., & Demir, E. (2018). Competencies of middle school teachers to prepare open-ended items used in open-ended tests for in-classroom assessment. *Elementary Education Online*, 17(4), 1912-1927. doi:10.17051/ilkonline.2019.506900
- Johnson, M. A. (2020). Years of experience and the ability to address student misconceptions in test items. *Mathematics Education Quarterly*, 50(4), 320-335.
- Kazuko, W. (2018). Assessing teachers' testing skills in ELT and enhancing their professional development through distance learning on the net. *Turkish Online Journal of Distance Education*, 5(1), 113-125.
- Kitchen, H., Bethell, G., Fordham, E., Henderson, K. & Ruochen Li, R. (2019). *OECD Reviews of Evaluation and Assessment in Education: Student Assessment in Turkey*. OECD Reviews of Evaluation and Assessment in Education. Paris: OECD Publishing. <https://doi.org/10.1787/5edc0abe-en>.
- Kumar, R. (2017). Verbal exposure for irrational fears: New directions in research and applications. *Clinical and Experimental Psychology*. doi:10.4172/2471-2701.1000157
- Kumar, R., Sameer, A., & Singh, B. (2012). Preliminary test of cognitive drill as an intervention. *Indian Journal of Clinical Psychology*, 39, 67-74.
- Loyalka, P., Popova, A., Li, G., & Shi, Z. (2018). Does teacher training actually work? Evidence from a large-scale randomized evaluation of a national teacher training program. *Rural Education Action Program (REAP), Working Paper 330*. Retrieved from: [https://fsi-live.s3.us-west-1.amazonaws.com/s3fs-public/does\\_teacher\\_training\\_work.pdf](https://fsi-live.s3.us-west-1.amazonaws.com/s3fs-public/does_teacher_training_work.pdf)
- Martinez, E. K. (2018). Reflective practice and test item writing: the role of experience. assessment in education: *Principles, Policy & Practice*, 35(1), 56-72.
- Nicol, C. (1999). Learning to teach mathematics: Questioning, listening, and responding. *Educational Studies in Mathematics*, 37, 45-66. <https://doi.org/10.1023/A:1003451423219>
- Nworgu, B. G. (2019). *Educational measurement and evaluation theory and practice* (Revised 2<sup>nd</sup> Edition). Nsukka: University Trust Publishers.
- Ololube, N. P. (2008). Evaluation competencies of professional and non-professional teachers in Nigeria. *Studies in Educational Evaluation*, 34(1), 44-51.
- OECD (2009). Creating effective teaching and learning environments: First results from TALIS Retrieved from: <http://www.oecd.org/education/school/43023606.pdf>
- OECD (2016). *PISA 2015 results (Vol I): Excellence and equity in education*. Paris: PISA, OECD Publishing. Retrieved from: <http://dx.doi.org/10.1787/9789264266490-en>
- OECD (2018). *Bridging the digital gender divide includes, upskill, innovative*. Retrieved from <https://www.oecd.org/digital/bridging-the-digital-gender-divide.pdf>
- OECD (2019). *TALIS 2018 results (Vol I): Teachers and school leaders as lifelong learners*, TALIS, Paris: OECD Publishing. Retrieved from: <https://doi.org/10.1787/1d0bc92a-en>.
- Paudel, T. (2020). Gender issue in teaching and learning mathematics. Conference Paper: Lecture at Tribhuvan University and M Phil Scholar of Kathmandu University.
- Quansah, F., Amoako, I., & Ankomah, F. (2019). Teachers' test construction skills in senior high schools in Ghana: Document analysis. *International Journal of Assessment Tools in Education*, 6(1), 1-8.

- Reardon, S. F., & Ho, A. D. (2015). Practical issues in estimating achievement gaps from coarsened data. *Journal of Educational and Behavioral Statistics*, 40(2), 158–189.
- Reardon, S. F., Kalogrides, D., Fahle, E. M., Podolsky, A. & Zárate, R. C. (2018). The relationship between test item format and gender achievement gaps on math and ELA tests in fourth and eight grades. *Educational Researcher*, 47(5), 284–294. doi: 10.3102/0013189X18762105
- Sahin, A. & Kulm, G. (2008). Sixth grades mathematics teachers' intentions and use of probing, guiding and factual questions. *Journal of Mathematics Teacher Education*, 11(3), 221-241. doi:10.1007/s10857-008-9071-2
- Şata, M. (2016). Türk Eğitim Sistemi'nde Sınıf İçi ile Geniş Ölçekli Ölçme ve Değerlendirmeye Genel bir Bakış. *Current Research in Education*, 2(1), 53-60.
- Scheiber, C., Reynolds, M. R., Hajovsky, D. B., & Kaufman, A. S. (2015). Gender differences in achievement in a large, nationally representative sample of children and adolescents. *Psychology in the Schools*, 52(4), 335-348. doi: 10.1002/pits.21827
- Sebastian M. A. (2020). Classification of test items written by mathematics pre-service teachers. *Philippines International Electronic Journal of Mathematics Education* 15 (2), em0577. <https://doi.org/10.29333/iejme/7599>
- Sharma, N. (2015). *Professional development of teacher*. New Delhi Publishers.
- Smith, J. (2019). The impact of teachers' experience on item construction expertise. *Educational Assessment*, 45(3), 210-225.
- Smith, L. B. (2020). Unpacking gender bias in mathematics teachers' test item writing. *Educational Psychology Review*, 42(3), 310-325.
- Stiggins, R.J., Griswold, M. M., & Wikelund, K. R. (1989). Measuring thinking skills through classroom assessment. *Journal of Educational Measurement*, 26, 233-246.
- Thompson, T. (2008). Mathematics teachers' interpretation of higher-order thinking in Bloom's taxonomy. *International Electronic Journal of Mathematics Education* 3(2), 96-109.
- Turner, S. L. (2019). Role of experienced teachers in nurturing item writing skills. *Professional Development in Education*, 47(3), 421-438.
- UNDP. (2006). *Nepal: Readings in human development*. Pulchowk, Kathmandu: United Nations Development Program.
- Verma, S., Kumar, R., Arya, B., & Kandhari, N. (2016). Treatment of multiple phobias through cognitive drill technique. *International Conference of Indian Academy of Health Psychology*. Gautam Buddha University, Greater Noida.
- Williams, R. D. (2021). Evolution of test item writing strategies with experience: Adapting to changing practices. *Teaching and Teacher Education*, 55, 102-115.