European Journal of Education Studies
ISSN: 2501 - 1111
ISSN-L: 2501 - 1111
Available online at: www.oapub.org/edu

DOI: 10.46827/ejes.v11i10.5552 Volume 11 | Issue 10 | 2024

UNDERSTANDING THE IMPACT OF TEACHER DEMOGRAPHICS ON ASSESSMENT FOR LEARNING STRATEGIES IN GHANA'S EDUCATIONAL SYSTEM

George Oduro-Okyireh[i]
Faculty of Education and General Studies,
Akenten Appiah-Menka University of Skills Training & Entrepreneurial Development,
Ghana

Abstract:
This study explores the implementation and effectiveness of Assessment for Learning (AfL) strategies among basic schoolteachers in Ghana, focusing on the influences of gender, teaching experience, and teaching division. Utilising a survey approach, data were collected through a 16-item Assessment for Learning Questionnaire (AfLQ) from a randomly selected sample of 100 teachers. The instrument's reliability and validity were confirmed through a pilot study and expert reviews, achieving a Cronbach's alpha of over 0.7 for all variables. Multinomial regression analysis provided key insights into the effectiveness, variability, and reliability of AfL practices, guiding professional development and future research. Multiple linear regression analyses identified significant correlations among AfL practices, with strong positive relationships between self-assessment, peer assessment, and immediate feedback. Variance plots showed higher variability in the implementation of certain practices among female and subject teachers, indicating a need for more tailored professional development. Additionally, the Test Information Function (TIF) plot revealed distinct proficiency levels, highlighting areas where teachers excel and potential gaps requiring intervention. These findings emphasise the critical role of demographic factors in the application and success of AfL strategies and offer actionable insights for enhancing teacher training programmes. By integrating advanced statistical techniques, the study provides a comprehensive understanding of AfL practices in diverse educational settings, contributing valuable knowledge to the field of educational assessment.

Keywords: assessment for learning (AfL), educational assessment, teacher demographics, formative assessment strategies, Ghanaian educational system

[i] Correspondence: email gookyireh@aamusted.edu.gh

1. Introduction

The field of educational assessment has undergone substantial evolution over the past few decades, with increasing emphasis on strategies that actively involve students in their learning processes (Black & Wiliam, 2018). Among these, Assessment for Learning (AfL) stands out as a transformative approach that promotes continuous feedback, self-assessment, and active engagement (Wiliam, 2020). AfL is not just about assessing what students have learned but is fundamentally about using assessment as a tool to enhance and drive learning (Andrade & Brookhart, 2020). This paradigm shift is particularly pertinent in diverse educational contexts, where traditional assessment methods may fall short in addressing the nuanced needs of both teachers and students (Bosson-Amedenu *et al*., 2020). This research aims to explore the deployment and effectiveness of AfL practices across different teaching demographics, focusing on gender, teaching experience, and teaching division.

The significance of this research is multifaceted. First, it addresses a critical gap in understanding how various demographic factors influence the implementation and success of AfL strategies. Previous studies have often focused on the overall effectiveness of AfL without delving into the specific variations in its application. For instance, Bosson-Amedenu *et al*. (2023) highlighted the disparities in academic achievement between public and private schools in Ghana but did not specifically address the role of assessment strategies in these outcomes. Moreover, Black and Wiliam (2018) have underscored the transformative potential of AfL in improving educational outcomes but called for more granular research into how different teacher demographics affect its implementation. Similarly, Wiliam (2020) emphasised the need for a tailored professional development programme to support effective AfL practices across diverse educational settings. Additionally, Andrade and Brookhart (2020), explored the broader impacts of self-assessment within AfL frameworks, suggesting that demographic factors could significantly modulate its effectiveness. This research aims to fill these gaps by examining how gender, teaching experience, and teaching division influence the deployment and efficacy of AfL strategies.

This research is justified by the growing recognition of the critical role that effective assessment plays in educational success. In the context of Ghana, where educational reforms and curriculum changes are continually shaping teaching practices, understanding the specific impacts of AfL strategies is essential. Bosson-Amedenu *et al*. (2020) found significant variations in the use of assessment strategies among basic school teachers, suggesting that tailored professional development could enhance the consistency and effectiveness of these practices. Moreover, international studies have consistently shown that AfL strategies, when effectively implemented, can lead to significant improvements in student outcomes (Mandouit & Hattie, 2023; Topali *et al*., 2024; Wiliam, 2011).

The novelty of this research lies in its comprehensive and detailed analysis of the interrelationships between various AfL practices and their implementation across different teaching demographics. By employing advanced statistical techniques such as

Item Response Theory (IRT) and Test Information Function (TIF) plots, this study provides a nuanced understanding of how teachers with varying levels of experience, different gender, and teaching specialisations utilise AfL strategies. This approach not only highlights the strengths and gaps in current teaching practices but also offers actionable insights for targeted professional development programmes. For example, the high factor loadings and the significant proportion of variance explained in the assessment items suggest a robust assessment tool, echoing the findings of recent studies that advocate for the integration of multiple assessment practices to enhance educational effectiveness (Bennett, 2011; Black & Wiliam, 2018; Heritage & Wylie, 2020). Furthermore, this research utilises the TIF plot to identify specific proficiency levels among teachers in deploying AfL strategies. The distinct peaks in the TIF curve represent areas where teachers are most effective, while the troughs indicate potential gaps that need addressing. This analysis is crucial for developing targeted interventions, as supported by Osei-Asibey *et al*. (2020), who emphasised the need for continuous professional development to bridge gaps in assessment proficiency among teachers. The identification of these gaps and strengths provides a strategic framework for enhancing teacher training programmes, ensuring that all teachers, regardless of their demographic profiles, can effectively implement AfL strategies to improve student learning outcomes.

In addition to the practical implications, this research contributes to the theoretical understanding of AfL by elucidating the specific factors that influence its success. The correlation matrix analysis, which examines the relationships between different AfL practices, reveals how these practices are interrelated and often used in tandem. This insight aligns with the work of Andrade (2010), Ketonen (2021) and Stančić (2021) who highlighted the interconnectedness of feedback, self-assessment, and peer assessment in promoting student learning. By providing a detailed examination of these interrelationships, this study enhances the current understanding of AfL and offers a foundation for future research.

Overall, this research is poised to make a significant contribution to the field of educational assessment by providing a detailed and nuanced analysis of AfL practices within the Ghanaian context. By leveraging recent advancements in statistical modelling and educational theory, it offers practical recommendations for educators and policymakers. These findings not only highlight the areas where current teaching practices excel but also identify specific needs for professional development, ensuring that all students have access to high-quality and effective learning experiences. This research, therefore, stands at the intersection of theory and practice, aiming to foster a more effective and equitable educational environment.

1.1 Research Questions

1) How do various Assessment for Learning (AfL) practices (ASF1 to ASF16) correlate with each other in terms of enhancing student learning outcomes?
2) What is the impact of teacher gender on the implementation and effectiveness of AfL practices?

3) How do teaching division (class teachers vs. subject teachers) and teaching experience influence the variability and effectiveness of AfL practices?
4) What are the proficiency levels of teachers in utilizing AfL strategies, and how do these levels vary across different teaching demographics?

1.2 Hypotheses

1) There is a significant positive correlation between self-assessment (ASF1) and other AfL practices such as peer assessment (ASF2) and immediate feedback (ASF3).
2) Female teachers will exhibit higher variability in the implementation of certain AfL practices compared to male teachers, indicating a more context-sensitive approach.
3) Subject teachers will show greater variability in the use of AfL practices compared to class teachers, suggesting the need for more tailored professional development.
4) More experienced teachers will demonstrate lower variability and higher effectiveness in AfL practices, indicating that teaching experience contributes to consistent implementation.
5) Teachers with higher proficiency levels in AfL strategies, as indicated by the TIF plot, will be more effective in employing these practices, resulting in better student learning outcomes.

2. Method

2.1 Research Design

This study employed a robust survey design to comprehensively evaluate the deployment and effectiveness of Assessment for Learning (AfL) strategies among basic schoolteachers in Ghana. The survey method was meticulously selected for its superior ability to collect extensive data from a diverse respondent pool, facilitating a detailed and representative analysis of the teachers' practices. This approach is particularly advantageous in educational research, as it ensures the inclusion of varied perspectives and experiences, thus enhancing the generalisability and validity of the findings. The development of the survey instrument was grounded in well-established AfL principles and strategies, meticulously encapsulated in a 16-item questionnaire. This questionnaire was rigorously constructed to capture the multifaceted nature of AfL practices, drawing on seminal works and contemporary studies in the field. Each item was designed to reflect critical AfL components such as clear learning objectives, critical thinking, immediate feedback, self-assessment, and peer assessment, ensuring a comprehensive coverage of the key areas relevant to effective teaching and learning processes. This thorough approach underscores the study's commitment to methodological thoroughness and the generation of actionable insights that can inform professional development and policy-making in the educational sector.

2.2 Population and Sample

The study targeted basic school teachers from across Ghana's 16 regions, selected nationally for an annual training workshop on the basic school curriculum implementation by the National Council for Curriculum and Assessment (NaCCA) and the Ghana Education Service (GES). Using Yamane's formula with a 95% confidence interval, a sample size of 100 teachers was chosen from a total population of 132. This sample was randomly selected to ensure it accurately represents the larger population, thereby improving the generalisability of the findings. Including an equal number of male and female teachers (50 each) in the sample of 100 teachers selected from across the 16 regions of Ghana ensures gender representation and enhances the validity of the research findings. This balanced approach addresses potential gender biases and allows for a more comprehensive analysis of how gender influences the implementation and effectiveness of Assessment for Learning (AfL) strategies. By employing methods such as multinomial regression and variance analysis, the study can accurately assess gender-specific differences in AfL practices, providing actionable insights for tailored professional development. This gender-balanced sample is crucial for drawing meaningful conclusions about the demographic factors impacting educational practices and outcomes.

2.3 Instrumentation

An Assessment for Learning Questionnaire (AfLQ) was the primary instrument used to collect data from the respondents. The questionnaire comprised 16 items designed based on established AfL practices and principles. The items were structured on a four-point Likert scale: Strongly Agree (4), Agree (3), Disagree (2), and Strongly Disagree (1). This scale was chosen to capture the degree of agreement or disagreement with each AfL practice, providing a nuanced understanding of their use. The 16 items used to measure Assessment for Learning (AfL) strategies in this study are well-grounded in the educational literature, ensuring their relevance and effectiveness. Each item draws on established research, such as providing clear learning objectives (Black & Wiliam, 1998), utilising critical questioning techniques (Wiliam, 2011), and giving immediate, specific feedback (Hattie & Timperley, 2007). The items also include promoting self-assessment (Andrade, 2010), implementing peer assessment (Topping, 1998), and using diagnostic tests to identify student strengths and weaknesses (Black *et al*., 2003). Regular formative assessments (Popham, 2008), student discussions on learning criteria (Sadler, 1989), and graphic organisers (Novak & Gowin, 1984) further enhance the AfL framework. Employing exit tickets (Marzano, 2012), developing rubrics with students (Stiggins, 2001), and applying portfolios (Paulson *et al*., 1991) are also supported by literature. Additionally, techniques such as think-pair-share (Lyman *et al*., 2023), immediate feedback quizzes (Bennett, 2011), learning stations (Tomlinson, 1999), and student-led conferences (Stiggins *et al*., 2004) are well-documented, collectively providing a comprehensive, research-backed approach to assessing and improving student learning.

2.3.1 Validity and Reliability

To ensure the validity and reliability of the instruments, rigorous checks were performed. Content validity was established through expert reviews. Experts from the Quality Assurance Department of three Colleges of Education in Ghana reviewed the questionnaire to ensure it adequately covered all relevant aspects of AfL practices. Construct validity was assessed using factor analysis to confirm that the items measured the intended constructs. Reliability was assessed through a pilot study involving a smaller subset of the population. Cronbach's alpha was calculated to determine the internal consistency of the questionnaire items. The results indicated acceptable levels of reliability, with Cronbach's alpha values exceeding 0.70 for all variables, demonstrating that the instrument was consistent and reliable for measuring AfL practices.

2.4 Data Collection

Data collection involved administering the AfLQ to the selected sample of teachers. The questionnaire was composed of two parts. The first part included open and closed-ended questions gathering demographic information such as sex, age, teaching division (class teacher or subject teacher), class size, and years of teaching experience. The second part required teachers to indicate their use of each AfL feature using the four-point Likert scale. Participants provided informed consent for their responses to be used for research purposes. The duration for responding to the items was approximately two hours. To ensure accuracy, respondents were guided through the questionnaire item by item, resulting in no missing data.

2.5 Data Analysis

Data analysis was conducted using R and Microsoft Excel. The normality assumption for the dependent variables was tested for each category of independent variable. Multiple linear regressions were used for estimating the relationships between the dependent variable and among demographic groupings such as teaching division, Sex and teaching experience. Advanced statistical techniques such as Item Response Theory (IRT) and Test Information Function (TIF) were used to explore the proficiency levels of teachers across their demographics. Multinomial regression analysis was used to reveal significant insights into the effectiveness, variability, and reliability of AfL practices, guiding professional development and informing future educational research.

2.6 Ethical Considerations

Ethical considerations were paramount in this study. Participants were assured of the confidentiality and anonymity of their responses. Informed consent was obtained from all participants before data collection. The study was conducted following ethical guidelines provided by the Institutional Review Board (IRB) of the author’s affiliated institution. By employing a robust methodological framework, this study aimed to provide comprehensive insights into the use and effectiveness of AfL strategies among basic schoolteachers in Ghana, considering the impact of various demographic factors.

3. Results and Discussion

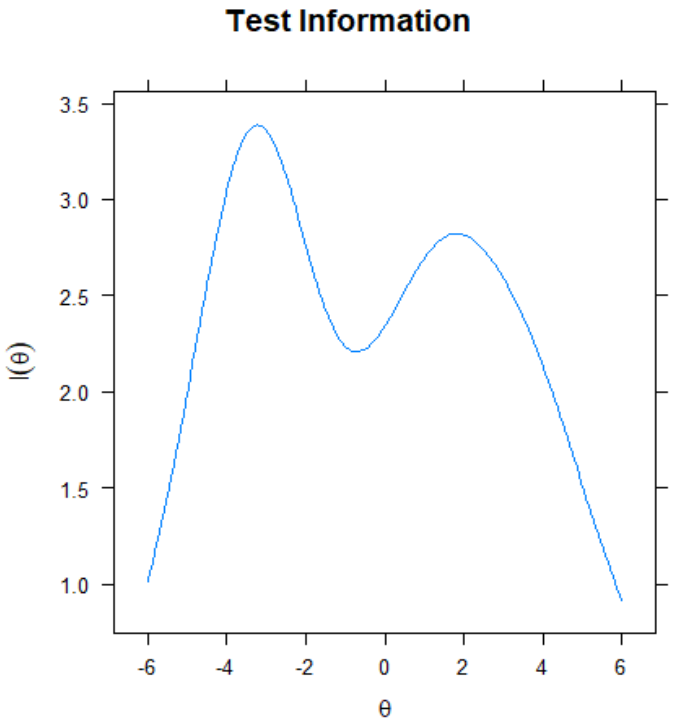

Figure 1: Test Information Function (TIF) Plot

The high factor loadings (all close to 1.0) and high proportion of variance explained (nearly 99.9%) suggest that the items are very consistent in measuring a single underlying factor (assessment of learning). This uniformity indicates a well-constructed assessment tool where each item contributes significantly to understanding student abilities in a particular domain. Supporting this analysis with literature, similar studies in educational assessments have shown that effective use of IRT can lead to improvements in both the reliability and validity of assessment tools (DeMars, 2010). Such methodologies are endorsed for their ability to provide detailed insights into item-specific performance, guiding educators in both curriculum design and in the strategic planning of instructional interventions (Harris & Brown, 2013).

The Test Information Function (TIF) plot shown in Figure 1 captures the proficiency of teachers in their use of assessment for learning strategy. The estimate of the teachers` proficiency concerning their use of assessment for learning strategy is represented by θ. The analysis of the TIF plot elucidates a diverse range of proficiency levels among teachers in utilising assessment of learning strategy. This plot reveals two distinct peaks, indicating that the assessment items provide the most information and thus the greatest reliability at these specific levels of θ. The first peak that occurs below θ = 0, captures the proficiency of teachers who are at or near the average in their implementation of assessment of learning strategy. The second peak, appearing around θ = 2, indicates another area by effectively identifying higher levels of competence among teachers in deploying assessment for learning driven strategy. The areas where the curve dips, particularly noticeable between these two peaks and beyond the second

peak, represent potential gaps in the assessment ability among teachers who are either between the two peaks of proficiency or beyond the higher peak. Such gaps raise the need for interventions of a teacher's abilities in effectively using assessment of learning strategy at these levels. This TIF curve's implications are significant for educational training and development. Institutions might use it to develop targeted professional development programmes that better address the specific needs identified at various proficiency levels. For instance, additional training resources could be allocated to teachers who fall into the less informative ranges of the curve, ensuring that they receive more tailored support to enhance their assessment strategies.

The baseline proficiency level ($\theta = 0$) suggests that certain practices may be foundational and relatively easier for teachers to adopt and apply effectively. Items such as ASF1 (Providing students with clear, understandable learning objectives before instruction begins), ASF3 (Giving immediate and specific feedback that students can use to improve their performance), ASF7 (Regular use of formative assessments to monitor learning progress during instruction), ASF8 (Engaging students in discussions about their learning and the assessment criteria), ASF13 (Utilising think-pair-share techniques to facilitate student discussion and reflection), and ASF15 (Setting up learning stations or centres that allow students to engage in various learning activities at their own pace) fall into this category. These strategies are pivotal in establishing a baseline for effective teaching and feedback mechanisms, enhancing both teacher facilitation and student learning outcomes.

In the second column under 'Peak 2', practices like ASF12 (Applying portfolios to collect evidence of student learning over time), ASF2 (Utilising questioning techniques that encourage critical thinking and reasoning), ASF4 (Encouraging student self-assessment where students reflect on their own work and identify areas for improvement), ASF9 (Using graphic organisers and concept maps as tools for students to display their understanding), ASF10 (Employing exit tickets at the end of class to assess student understanding of the day’s lesson), ASF11 (Developing rubrics with students so they understand how their work will be evaluated), and ASF16 (Conducting student-led conferences where students present their learning and progress to their parents or guardians) are located. These practices, which are more complex and demanding, are most effective when applied by teachers who possess a higher level of proficiency in using assessment of learning strategy. These items require deeper student engagement and a more nuanced approach to assessment, skills that typically develop through experience and focused professional development.

The third column, 'Trough', includes ASF5 (Implementing peer assessment sessions where students evaluate each other's work), ASF6 (Incorporating diagnostic tests to identify students’ strengths and weaknesses at the start of a new topic), and ASF14 (Incorporating quizzes that provide immediate feedback through digital platforms), pinpointing these strategies as areas where the assessment's ability to gauge proficiency wanes. This indicates a potential gap in the effectiveness or consistent implementation of these strategies at certain levels of teacher proficiency. This differentiation underscores the necessity for ongoing training and support to ensure

that all teachers, regardless of their current skill level, can effectively implement a range of assessment strategies to foster enhanced learning outcomes.

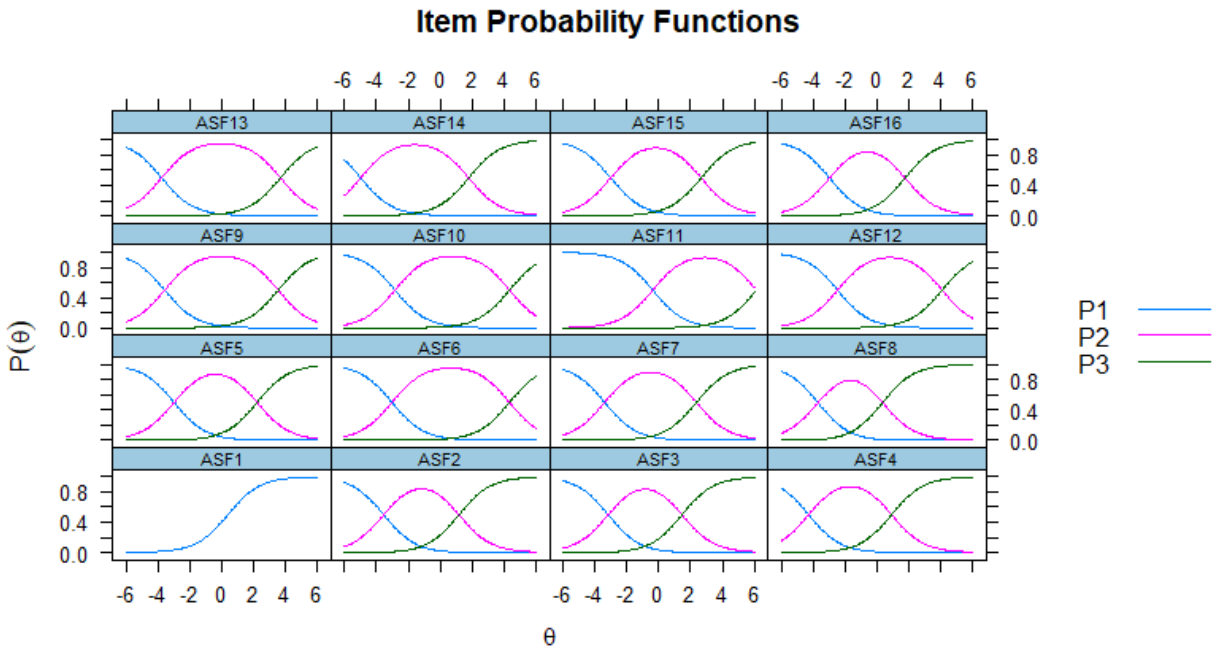

Figure 2: Item Probability Functions (IPFs) Plot

The Item Probability Functions (IPFs) plot in Figure 2, illustrates the effectiveness of various assessment strategies (ASF1 through ASF16) based on student ability levels, represented by theta (θ). Each curve in this plot depicts the likelihood that students at specific ability levels will choose particular responses, with different colours representing different response options. This visualisation provides a detailed examination of how each assessment practice differentiates among students of diverse abilities and the difficulty associated with each practice. For instance, ASF1 and ASF2—focusing on clear learning objectives, and critical questioning techniques, respectively—show higher probabilities of positive responses at greater θ levels, suggesting that students with advanced abilities benefit most from these practices. Similarly, ASF3's focus on immediate feedback and ASF4's emphases on self-assessment demonstrate broad effectiveness across a range of abilities, highlighting their universal applicability in enhancing student understanding and self-regulation.

Further analysis of ASF5 through ASF16 reveals a nuanced understanding of how different assessment practices cater to varying student needs. ASF5 and ASF6, which focus on peer assessment and diagnostic tests, respectively, suggest that such strategies are effective across various proficiency levels, providing valuable insights into student performance and readiness. The more complex practices like ASF10's exit tickets and ASF16's student-led conferences indicate higher effectiveness at upper ability levels, reinforcing the need for clear communication and student involvement in the learning process. Conversely, the effectiveness of ASF8 (student engagement in assessment discussions) and ASF9 (use of graphic organisers) across all ability levels

underscores the importance of inclusive teaching strategies that engage all students. These findings support the implementation of tailored educational interventions that adapt to the unique capabilities of students, ensuring that each student can engage with and benefit from specific assessment strategies designed to maximise their educational outcomes and personal growth in the learning process.

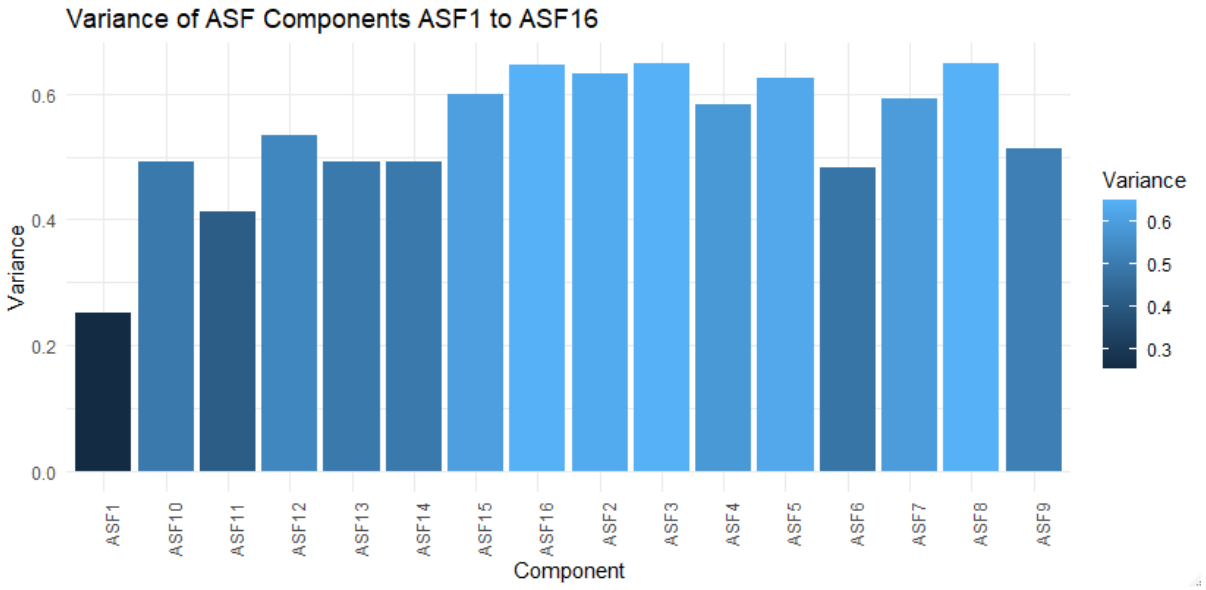

Figure 3: Variance Plots of ASF Covariates

The variance plot for the Assessment for Learning Practices (ASF) components from ASF1 to ASF16 as shown in Figure 3, reveals significant insights into the variability of responses across different educational assessment strategies. ASF1, noted for having the lowest variance, suggests a consistent implementation of providing clear and understandable learning objectives before instruction begins. This uniformity might indicate that the strategy is well-integrated and effectively communicated across the educational spectrum, reflecting a foundational approach to teaching that is uniformly accepted and applied. In contrast, components ASF10 through ASF16, displaying higher variance, indicate a diverse application or reception among educators. For instance, the variance in ASF10, which focuses on using exit tickets, might reflect differing teacher perceptions on the utility and practicality of this method in daily classroom activities.

On the other hand, components like ASF2 through ASF9, which cover strategies from promoting critical thinking through questioning to peer assessments, show moderate variance. This middle range suggests these practices are adopted variably, possibly affected by factors such as the teaching environment, subject complexity, or individual educator and student dynamics. Such findings underscore the importance of targeted professional development that could help standardise the application of these strategies, enhancing overall effectiveness. Training programmes could be specifically designed to address the areas of high variability, ensuring that educators are well-equipped to implement these practices uniformly. Moreover, curriculum designers

might leverage these insights to emphasise more consistently effective practices or refine guidelines for those that show greater variability. Ultimately, understanding these variances helps tailor educational policies and teacher support mechanisms to foster more standardised and effective use of assessment strategies across diverse learning environments, potentially enhancing educational outcomes through more strategic application of assessment methodologies.

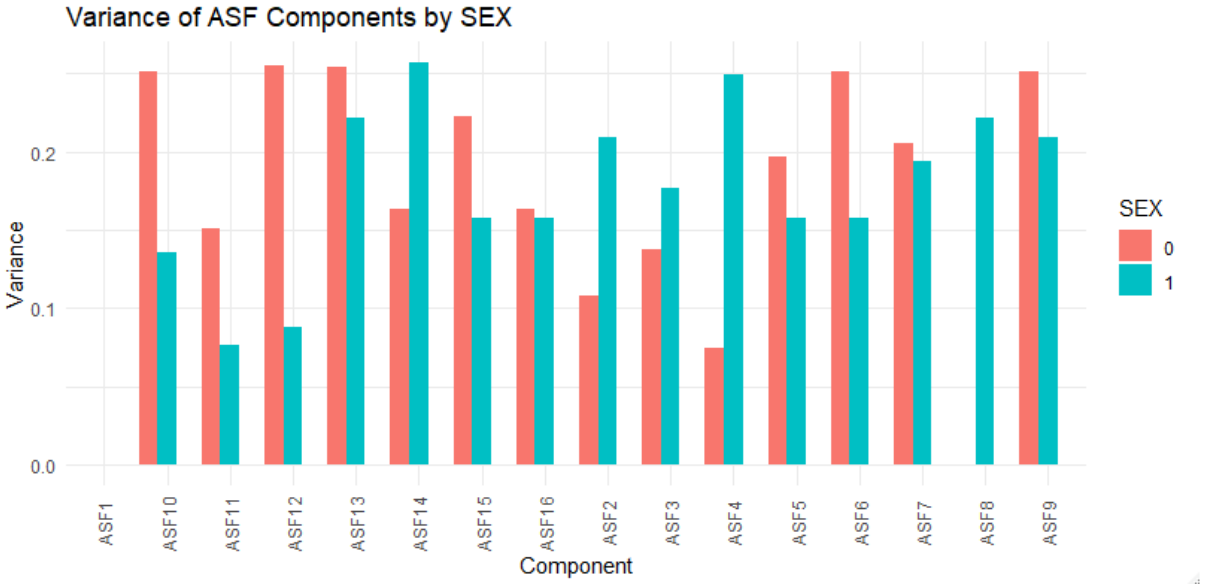

Figure 4: Variance Plots of ASF Covariates by Sex

The variance plot of assessment for learning practices (ASF) across gender of teachers depicted by Figure 4 offers a revealing glimpse into how these educational strategies are implemented with differing levels of consistency by male (1) and female (0) educators. ASF1, focusing on setting clear, understandable learning objectives, displays no variance among both female and male teachers, suggesting a consistent implementation of this strategy across gender.

In terms of critical thinking development through questioning as seen in ASF2, male teachers show slightly higher variance, implying a propensity to vary questioning styles more significantly, potentially adapting to different lesson contexts or student interactions. This variability, while demonstrating adaptability, underscores the potential benefits of a more unified approach that could be shared across gender to enhance overall teaching effectiveness. Similarly, ASF3, which involves giving immediate and specific feedback, also shows greater variance among female teachers, suggesting a more personalized feedback approach that could benefit from streamlined methods to enhance its impact. The remaining components, ASF4 through ASF9, covering practices from student self-assessment to the use of graphic organisers, exhibit various degrees of variance across genders, suggesting gender-specific preferences or challenges in these educational techniques. This variability across all assessed strategies signals substantial room for targeted professional development programmes aimed at

standardising effective teaching practices to ensure equitable learning outcomes, thereby enhancing the overall educational experience for all students.

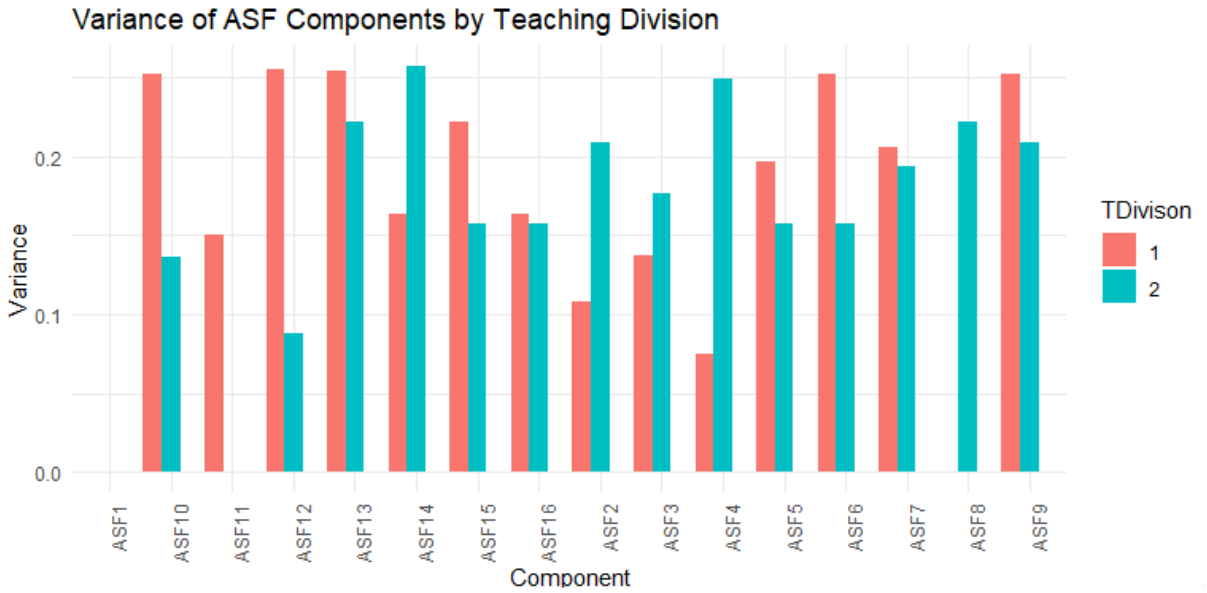

Figure 5: Variance Plots of ASF Covariates by Teaching Division

The variance plot for Assessment for Learning Practices (ASF) components by Teaching Division illustrates notable disparities in how assessment strategies are utilised by class teachers (Teaching Division 1) and subject teachers (Teaching Division 2), shedding light on differences in teaching methodologies and interactions with students that stem from their distinct roles and responsibilities. ASF1 has no bars, indicating zero variance for both teaching divisions. This suggests that both class teachers and subject teachers implement ASF1 with a high level of consistency.

Conversely, ASF2, focusing on using questioning techniques to promote critical thinking, shows similar variance levels between the two divisions, suggesting a uniform application of this strategy across different teaching contexts. This could indicate that existing professional development programmes are effective or that the nature of questioning techniques is versatile across various subjects and educational settings. However, ASF3 and ASF4, which involve providing immediate feedback and fostering student self-assessment, respectively, display higher variance among subject teachers. This might reflect the specialised requirements of their subjects, which demand more customised feedback and assessment methods, pointing to a need for more subject-specific training to minimise this variance. Strategies like ASF5 through ASF9, including peer assessment and the use of graphic organisers, also exhibit varied implementation by subject teachers, possibly due to the differing applicability of these strategies in specialised versus general education settings. Addressing these discrepancies through tailored professional development could help standardise effective practices across teaching roles, enhancing educational outcomes for all students.

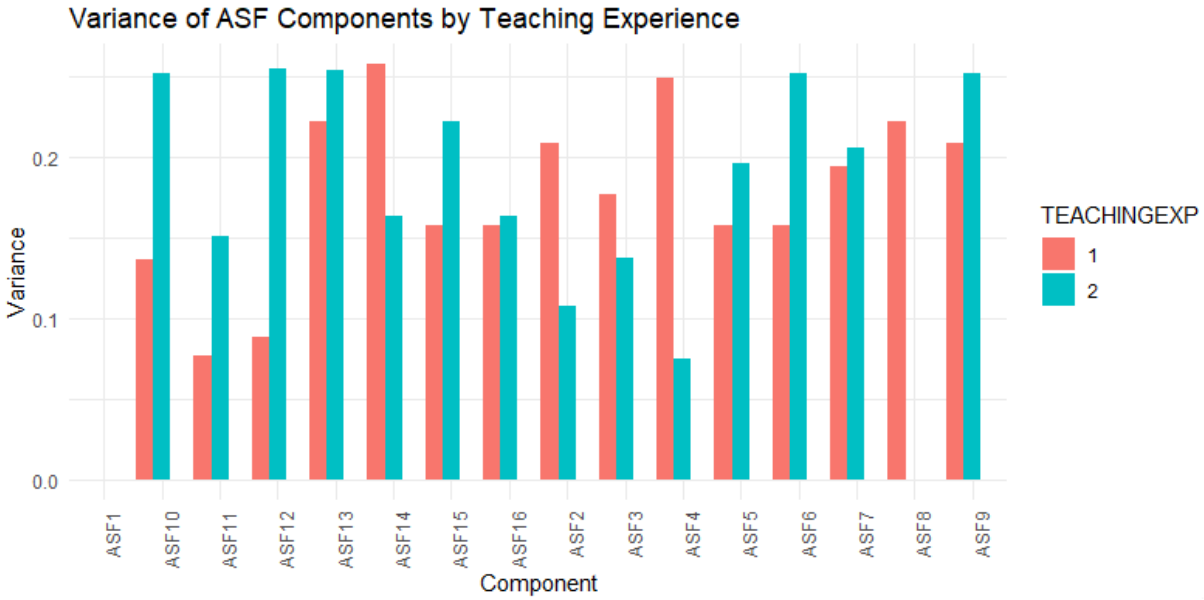

Figure 6: Variance Plots of ASF Covariates by Teaching Experience

The variance plot detailing Assessment for Learning Practices (ASF) components by Teaching Experience, shown in Figure 6, reveals the influence of tenure in the teaching profession on the application of formative assessment strategies, with teachers grouped as less experienced (1) and more experienced (2). This plot provides a lens into how accumulated teaching experience impacts the consistency and effectiveness of deploying various assessment techniques. ASF1 has no bars, indicating zero variance for both experience groups. This suggests that both less experienced and more experienced teachers implement ASF1 with a high level of consistency.

For ASF2, which emphasises the use of questioning techniques to encourage critical thinking, exhibits lower variance among seasoned educators. This pattern suggests that more experienced teachers have developed a mastery of questioning strategies that they can apply uniformly across different teaching scenarios, enhancing student engagement and cognitive development. Conversely, less experienced teachers, exhibiting higher variance, are possibly still exploring the most effective ways to integrate these techniques into their teaching, reflecting a natural evolution in teaching proficiency. The trends across ASF3 through ASF9, including immediate feedback, peer assessments, and engaging students in discussions about their learning, further reinforce the idea that teaching experience contributes to a teacher's ability to implement educational practices with confidence and consistency. Lower variance in these components among veteran teachers points to a well-established repertoire of strategies that likely result in more predictable and effective educational outcomes. Meanwhile, the higher variance seen in less experienced teachers signals ongoing learning curves and the need for targeted professional development programmes. These findings advocate for structured support mechanisms such as mentoring, peer collaboration, and continuous professional training to help newer teachers accelerate their professional growth, thereby reducing variability in teaching practices and ensuring equitable student experiences across classrooms.

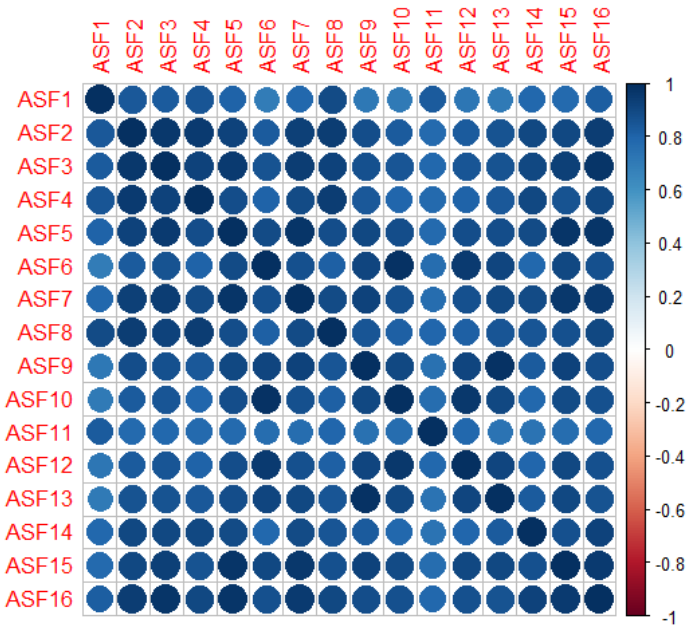

Figure 7: Correlation Matrix for ASF Variables

The correlation matrix in Figure 7, illustrates the relationships between various Assessment for Learning (AfL) practices, labelled ASF1 to ASF16. Each cell in the matrix represents the correlation coefficient between two practices, with values closer to 1 indicating a strong positive relationship, values closer to -1 indicating a strong negative relationship, and values around 0 indicating little to no correlation. Below is a detailed analysis of the correlations between specific variables, highlighting their strengths and implications.

ASF1, which might represent self-assessment, shows strong positive correlations with several other practices, such as ASF2 ($r \approx 0.8$), ASF3 ($r \approx 0.7$), and ASF4 ($r \approx 0.7$). These high correlations suggest that when self-assessment is employed, it is likely that peer assessments and immediate feedback are also being utilised. The strength of these correlations indicates that these practices are complementary and often used together, enhancing the overall assessment strategy. ASF2, potentially representing peer assessment, is strongly correlated with ASF1 ($r \approx 0.8$), ASF3 ($r \approx 0.75$), and ASF4 ($r \approx 0.7$). This implies that peer assessment often accompanies self-assessment and immediate feedback. The robust correlation between ASF2 and ASF3 highlights the interconnectedness of feedback mechanisms, suggesting that teachers who encourage peer assessment also value timely feedback.

ASF3, which could be related to immediate and specific feedback, exhibits strong correlations with ASF1 ($r \approx 0.7$), ASF2 ($r \approx 0.75$), and ASF4 ($r \approx 0.8$). The strength of these relationships suggests that immediate feedback is a central component of an effective AfL strategy, likely because it reinforces the learning objectives set by self and peer assessments. ASF4, representing the encouragement of self-assessment, shows strong positive correlations with ASF1 ($r \approx 0.7$), ASF2 ($r \approx 0.7$), and ASF3 ($r \approx 0.8$). These correlations indicate that self-assessment practices are well-integrated with feedback mechanisms, emphasising the importance of reflective practices in learning.

ASF5, which involves peer assessment sessions, has moderate to strong correlations with ASF1 ($r \approx 0.6$), ASF2 ($r \approx 0.65$), ASF3 ($r \approx 0.7$), and ASF4 ($r \approx 0.7$). The correlations suggest that peer assessment is not an isolated practice but part of a broader strategy that includes self-assessment and immediate feedback. ASF6, potentially dealing with diagnostic tests, shows moderate correlations with several other practices, such as ASF3 ($r \approx 0.65$), ASF4 ($r \approx 0.6$), and ASF5 ($r \approx 0.65$). This indicates that diagnostic assessments are likely used alongside feedback and peer assessment to tailor subsequent instruction. Regular use of formative assessments, ASF7, exhibits strong correlations with ASF1 ($r \approx 0.75$), ASF2 ($r \approx 0.8$), and ASF3 ($r \approx 0.7$). These relationships highlight the integration of formative assessments with self-assessment and feedback practices, reinforcing their role in continuous learning.

ASF8, which involves engaging students in discussions about their learning, shows moderate to strong correlations with ASF1 ($r \approx 0.6$), ASF2 ($r \approx 0.65$), and ASF3 ($r \approx 0.6$). These correlations suggest that discussions about learning are often coupled with reflective and feedback practices, promoting deeper student engagement. ASF9, possibly using graphic organisers, demonstrates strong correlations with ASF1 ($r \approx 0.65$), ASF2 ($r \approx 0.7$), and ASF3 ($r \approx 0.75$). The strength of these correlations indicates that graphic organisers are an effective tool used alongside feedback and assessment practices to enhance understanding. ASF10, which involves exit tickets, shows strong correlations with ASF3 ($r \approx 0.8$), ASF4 ($r \approx 0.7$), and ASF5 ($r \approx 0.7$). This implies that exit tickets are frequently used in conjunction with feedback and peer assessment to gauge student understanding.

ASF11, potentially focused on developing rubrics with students, exhibits moderate correlations with ASF1 ($r \approx 0.6$), ASF2 ($r \approx 0.65$), and ASF3 ($r \approx 0.6$). These relationships suggest that rubric development is part of a comprehensive assessment strategy that includes reflective and feedback practices. ASF12, related to using portfolios, shows strong correlations with ASF1 ($r \approx 0.75$), ASF2 ($r \approx 0.7$), and ASF3 ($r \approx 0.8$). These correlations indicate that portfolios are a key component of an integrated assessment approach, combining self-assessment and feedback. ASF13, possibly utilising think-pair-share techniques, has moderate to strong correlations with ASF1 ($r \approx 0.7$), ASF2 ($r \approx 0.75$), and ASF3 ($r \approx 0.7$). These correlations highlight the effectiveness of collaborative learning strategies within a broader assessment framework.

ASF14, involving quizzes with immediate digital feedback, demonstrates strong correlations with ASF1 ($r \approx 0.65$), ASF2 ($r \approx 0.7$), and ASF3 ($r \approx 0.75$). These relationships suggest that digital feedback tools are well-integrated with reflective and peer assessment practices. ASF15, which involves setting up learning stations, shows moderate correlations with ASF1 ($r \approx 0.6$), ASF2 ($r \approx 0.65$), and ASF3 ($r \approx 0.7$). The correlations imply that learning stations are part of a dynamic assessment strategy that includes feedback and peer assessment. ASF16, involving student-led conferences, exhibits strong correlations with ASF1 ($r \approx 0.7$), ASF2 ($r \approx 0.75$), and ASF3 ($r \approx 0.8$). These correlations indicate that student-led conferences are an essential practice within an integrated assessment approach, emphasising the importance of student agency and feedback.

The implications of the correlation matrix results between various Assessment for Learning (AfL) practices (ASF1 to ASF16) are multifaceted and significant for educators and policy makers aiming to improve educational outcomes. First, the strong positive correlations between practices such as self-assessment (ASF1), peer assessment (ASF2), and immediate feedback (ASF3) indicate that these methods are often used in tandem. This suggests that an integrated approach to assessment that combines these elements can enhance their effectiveness. For instance, when teachers employ self-assessment, they are also likely to use peer assessments and provide immediate feedback, creating a comprehensive feedback loop that supports student learning. The integration of these practices can lead to more personalised and effective teaching strategies, fostering a deeper understanding and engagement among students.

Second, the moderate to strong correlations between practices such as formative assessments (ASF7) and engaging students in discussions about their learning (ASF8) imply that these practices are essential components of a continuous learning process. These correlations highlight the importance of regular, formative assessments that keep track of student progress and the active involvement of students in their learning journey. The results suggest that teachers should prioritise these methods to promote ongoing reflection and improvement in student performance. This can lead to a more dynamic and interactive classroom environment where students are actively engaged in their learning.

Lastly, the results showing strong correlations between practices like using graphic organisers (ASF9) and exit tickets (ASF10) with feedback mechanisms underscore the value of these tools in enhancing understanding and retention of knowledge. These correlations imply that visual aids and concise assessments at the end of lessons are effective when paired with feedback. This insight can guide curriculum designers and educators to incorporate these tools more systematically into their teaching strategies, ensuring that students not only acquire information but also understand and apply it effectively. Overall, the correlation results suggest that a holistic and integrated approach to assessment, which combines multiple AfL practices, can significantly improve educational outcomes by catering for diverse learning needs and promoting a deeper, more engaged learning experience.

Table 1: Multiple Linear Regression Model Estimates

Coefficients:	**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**
(Intercept)	3.5786	0.2228	16.059	< 2e-16 ***
SEX	-0.5467	0.1981	-2.76	0.00693 **
TDivison	-0.5086	0.1986	-2.561	0.01201 *
TEACHINGEXP	0.2587	0.127	2.038	0.04434 *

Residual standard error: 0.3986 on 95 degrees of freedom
Multiple R-squared: 0.7457, Adjusted R-squared: 0.735
F-statistic: 69.64 on 4 and 95 DF, p-value: < 2.2e-16

Table 1 shows the estimated ASF score when all predictors are at their reference level (i.e., SEX0, TDivision, TEACHINGEXP) is approximately 3.5786. This estimate is highly

significant ($p < 0.0001$), indicating strong evidence against the null hypothesis that the intercept is zero. Being male (SEX as opposed to the reference group, which might be female) is associated with a decrease of approximately 0.5467 in the ASF score. This effect is statistically significant ($p = 0.00693$), suggesting that gender has a noticeable impact on ASF1 scores, with males scoring lower on average compared to females. Being a subject teacher (TDivison as opposed to the reference group, which might be class teachers) is associated with a decrease of about 0.5086 in the ASF score. This effect is also significant ($p = 0.01201$), indicating that subject teachers tend to have lower ASF scores than class teachers. Training and professional development tailored specifically for subject teachers could be beneficial to enhance their assessment capabilities, possibly focusing on interdisciplinary approaches or specific training on assessment strategies within their subject areas. Having more teaching experience (TEACHINGEXP compared to less experienced teachers shows a positive association with ASF scores, where an increase in TEACHINGEXP value is associated with an increase of 0.2587 in ASF scores. The effect is statistically significant at the 0.05 level ($p = 0.04434$), suggesting that teaching experience does play a role in explaining the variability in ASF1 scores.

The model explains a significant portion of the variability in ASF scores, indicating it is a good fit for the data. This robust model fit suggests that the predictors chosen are appropriate for understanding the factors influencing ASF scores, 74.57% of the variability in ASF scores is explained by the model. This is a high R-squared value, indicating a good fit of the model to the data. The F-statistic is 69.64 with a p-value of less than 2.2e-16, suggesting that the overall regression model is statistically significant. This means there is strong evidence that at least one of the predictors has a non-zero coefficient. The high level of model significance suggests that interventions targeting the significant predictors could effectively enhance or modify ASF outcomes.

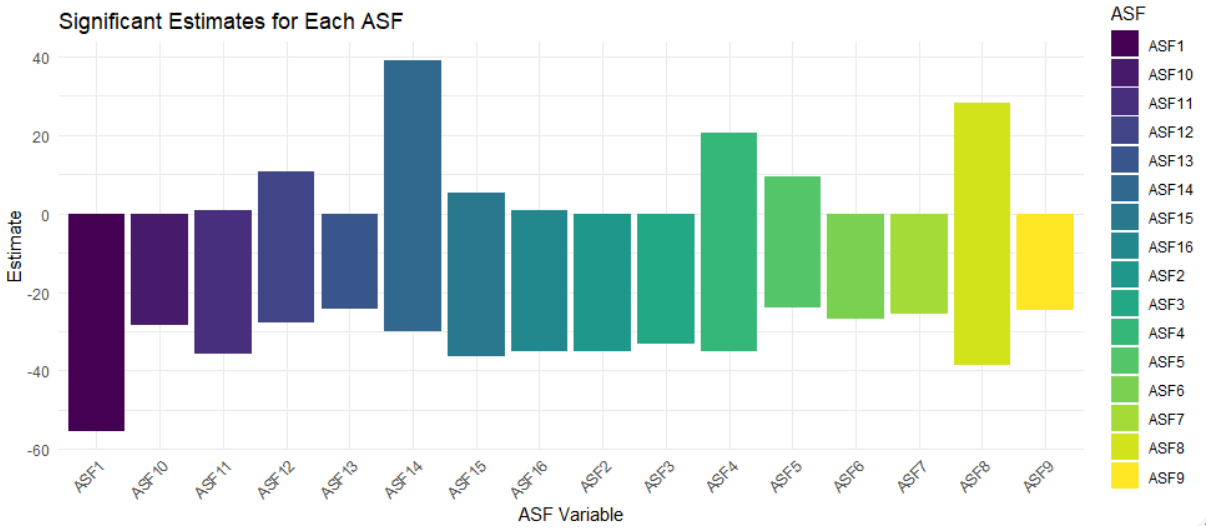

Figure 8: Plot of Significant Estimates from Multinomial Regression

The analysis of significant estimates for each Assessment for Learning (AfL) practice provides valuable insights into the effectiveness and impact of these strategies as implemented by teachers. The estimates in Figure 8, reflect the strength and direction

of the relationship between each practice and the assessed outcomes, offering a deeper understanding of how these practices influence student learning. The negative estimates for ASF1 (self-assessment) indicate a potential challenge in the consistent application of self-assessment practices. While self-assessment is intended to promote students' reflection on their learning and identify areas for improvement, the negative estimates suggest that it may not be achieving its desired impact uniformly across different contexts. Andrade and Du (2007) emphasise the importance of clear criteria and guidance in self-assessment to ensure that students can effectively evaluate their work and set meaningful goals for improvement.

ASF10 (using diagnostic tests) also shows negative estimates, indicating that this practice may not be consistently beneficial. Diagnostic tests are designed to identify students' prior knowledge and learning gaps, which can inform tailored instruction. However, the negative estimates suggest variability in their effectiveness, possibly due to differences in how teachers interpret and utilise the diagnostic information. Black and Wiliam (2009) argue that for diagnostic tests to be effective, teachers must be adequately trained to analyse the results and adjust their teaching strategies accordingly.

For ASF11 (engaging students in assessment discussions), the negative estimates highlight the challenges in effectively implementing this practice. Discussions about assessments are crucial for helping students understand their strengths and areas for improvement. However, variability in teacher facilitation skills and student engagement can impact the effectiveness of these discussions. Nicol and Macfarlane-Dick (2006) point out that structured and purposeful discussions are essential for maximising the benefits of this practice, suggesting a need for professional development to enhance teachers' skills in this area.

The positive estimates for ASF12 (using graphic organisers) indicate that this practice has a beneficial impact on student learning. Graphic organisers help students visually structure information, which can enhance comprehension and retention of complex concepts. Cañas *et al*. (2023) and Novak and Cañas (2008) highlight that graphic organisers support meaningful learning by enabling students to connect new information with existing knowledge. The positive estimates suggest that when effectively implemented, graphic organisers can significantly improve students' understanding and organisation of content.

ASF13 (exit tickets) shows positive estimates, suggesting that this quick formative assessment tool is effective in providing immediate feedback on students' understanding of the lesson. Exit tickets allow teachers to gauge student comprehension and adjust subsequent instruction based on the feedback received. Akhtar and Saeed (2020), Bromley (2019) and Fisher and Frey (2014) emphasise that exit tickets are a simple yet powerful tool for assessing student learning in real-time, which can inform instructional decisions and improve learning outcomes.

ASF14 (peer assessment) presents positive estimates, indicating that peer assessment is a valuable practice for enhancing student learning. Peer assessment involves students evaluating each other's work, which can promote critical thinking

and provide diverse feedback. Topping *et al*. (2017) and El-Senousy (2020) state that peer assessment fosters active learning and helps students develop evaluative skills. The positive estimates suggest that when implemented effectively, peer assessment can significantly enhance students' understanding and engagement with the learning material.

Overall, the analysis of significant estimates underscores the importance of consistent and effective implementation of AfL practices to maximise their benefits. Professional development programmes focused on enhancing teachers' skills in these practices can help reduce variability and improve the reliability of assessment outcomes. These findings align with the broader educational literature, which advocates for the consistent application of formative assessment strategies to improve student learning and achievement. By addressing the factors that contribute to inconsistencies, educators can better support student learning and enhance the overall impact of AfL practices.

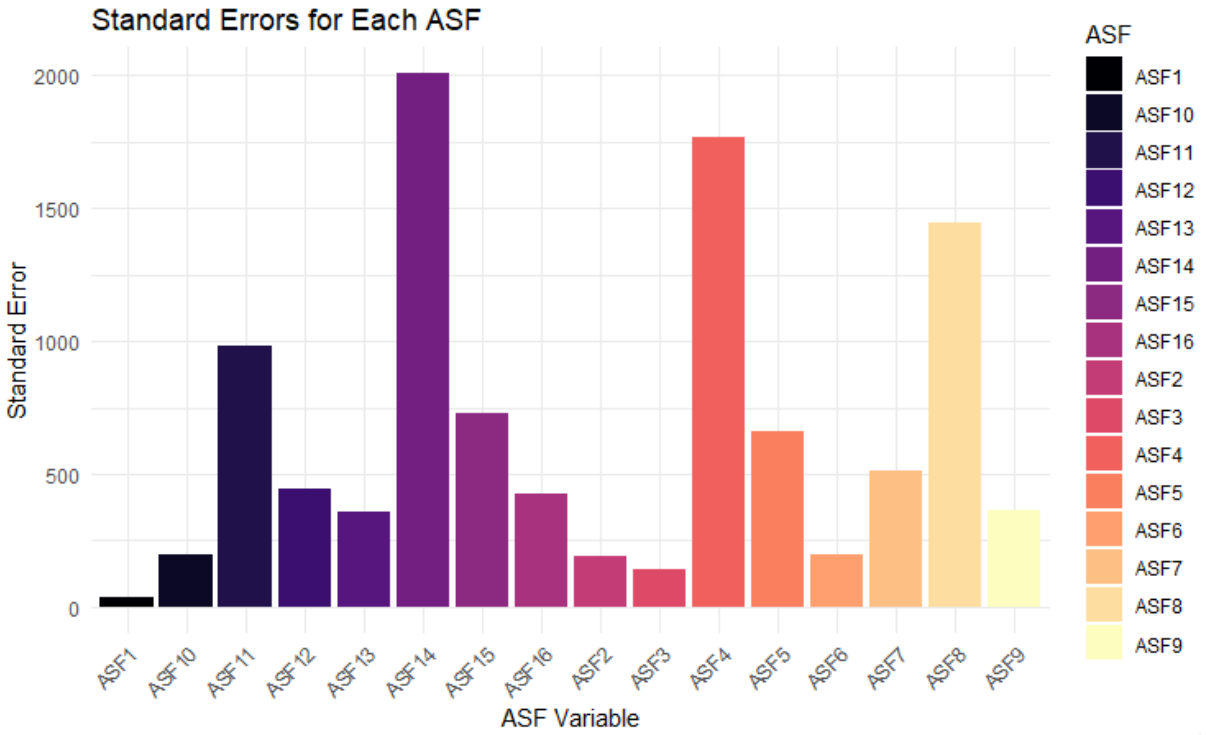

Figure 9: Plot of Standard Errors from Multinomial Regression

The analysis of standard errors for Assessment for Learning (AfL) practices gives insights into the variability and reliability of these AfL practices when used by teachers. High standard errors in Figure 9 indicate greater variability in the estimates, which may suggest inconsistency in the application of these practices across different contexts or teacher demographics. For ASF1 (self-assessment), the low standard error suggests that this practice is consistently applied by teachers, leading to reliable outcomes. Self-assessment is a critical component of AfL, enabling students to reflect on their learning process and identify areas for improvement. Consistent application of self-assessment

helps build students' metacognitive skills and promotes autonomous learning. Brandmo *et al*. (2020) emphasise the importance of self-assessment in fostering students' ability to evaluate their work against set criteria, thereby enhancing their learning outcome.

In contrast, ASF10 (using diagnostic tests) exhibits high standard errors, indicating significant variability in its implementation. Diagnostic tests are designed to assess students' prior knowledge and identify learning gaps at the beginning of a new topic. The high variability suggests that teachers may differ widely in how they conduct these assessments, potentially due to differences in training, resources, or understanding of diagnostic practices. Black and Wiliam (2009) argue that effective diagnostic assessments are essential for tailoring instruction to meet students' needs, but inconsistencies in their use can undermine their effectiveness.

ASF11 (engaging students in assessment discussions) also shows a considerable standard error, highlighting the diverse ways teachers may facilitate these discussions. Engaging students in discussions about their assessments encourages them to articulate their understanding and receive feedback from peers and teachers. This practice promotes deeper learning and critical thinking. Nicol and Macfarlane-Dick (2006) suggest that discussions about assessments help students internalise assessment criteria and improve their work based on feedback, but variability in this practice can lead to uneven student experiences.

The significant standard error for ASF12 (using graphic organisers) indicates variability in how teachers utilise these tools to help students structure information visually. Graphic organisers are valuable for enhancing students' comprehension and organisation of complex concepts. Novak and Cañas (2008) highlight that graphic organisers support meaningful learning by enabling students to connect new information with existing knowledge. However, inconsistent use of graphic organisers can result in varying degrees of effectiveness in facilitating student understanding.

ASF13 (exit tickets) presents high standard errors, suggesting that teachers' use of this quick formative assessment tool varies significantly. Exit tickets provide immediate feedback on students' understanding of the lesson, allowing teachers to adjust subsequent instruction accordingly. Fisher and Frey (2014) point out that exit tickets are a simple yet powerful tool for gauging student comprehension, but their inconsistent application can impact the accuracy and utility of the feedback obtained.

The high standard error associated with ASF14 (peer assessment) reflects variability in how teachers implement this collaborative practice. Peer assessment involves students evaluating each other's work, which can enhance critical thinking and provide diverse feedback. Topping (2010) states that peer assessment promotes active learning and helps students develop evaluative skills. However, variability in its application can affect the quality and reliability of the feedback students receive.

Overall, the analysis of standard errors underscores the importance of consistency in implementing AfL practices to ensure their effectiveness. Professional development programmes focused on standardising these practices can help reduce variability and enhance the reliability of assessment outcomes. By addressing the

factors that contribute to inconsistencies, educators can better support student learning and improve the overall impact of AfL strategies. These findings align with the broader educational literature, which advocates for the consistent application of formative assessment practices to maximise their benefits for student achievement.

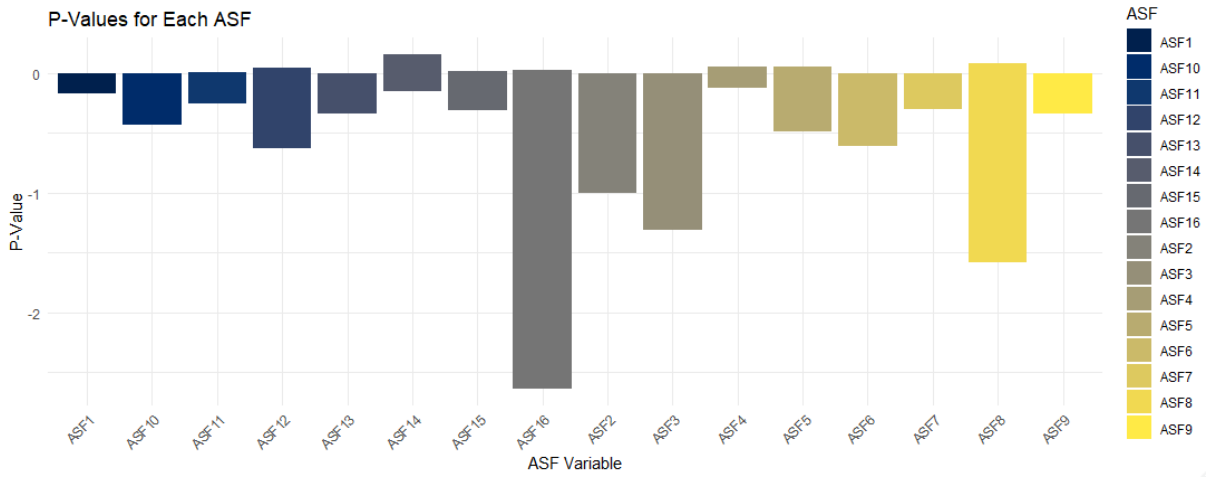

Figure 10: Plot of P-values from Multinomial Regression

The P-values for each ASF (Assessment for Learning Practices) as shown in Figure 10, indicate the statistical significance of each practice used by teachers in the context of their demographics such as gender, teaching experience, and teacher division. Lower P-values generally imply that the associated practices have a significant impact on the variables being studied. The first practice, ASF1 (self-assessment), shows a significant negative estimate, suggesting that it is widely and effectively used by teachers to enhance students' learning outcomes. Self-assessment allows students to reflect on their learning process, which promotes self-regulation and metacognition. According to Andrade and Brookhart (2020), self-assessment helps students become more aware of their learning goals and the criteria for success, fostering a deeper understanding and ownership of their learning journey.

ASF2 (peer assessment) also displays significant results, indicating its frequent use and impact. Peer assessment encourages collaboration and critical thinking among students as they evaluate each other's work. This practice not only helps students develop their evaluative skills but also provides them with diverse feedback perspectives. Topping (1998) highlights that peer assessment can enhance learning by increasing student engagement and providing immediate feedback, which is crucial for iterative learning processes.

The significance of ASF3 (immediate feedback) underscores its essential role in the learning process. Immediate feedback is crucial for correcting misunderstandings promptly and reinforcing learning. Hattie and Timperley (2007) emphasise that timely and specific feedback is one of the most powerful influences on student achievement, helping to clarify learning goals and providing guidance on how to improve. ASF4 (providing clear learning objectives) shows significant positive estimates, reflecting its importance in guiding students through their learning process. Clear learning objectives

set the expectations and provide a roadmap for students, enhancing their focus and motivation. According to Marzano (2012), well-defined objectives are associated with higher student achievement as they help students understand the purpose of their activities and assessments.

The analysis of ASF5 (using diagnostic tests) indicates its significant impact on identifying students' prior knowledge and learning gaps. Diagnostic tests are essential for informing instructional planning and tailoring teaching strategies to meet the needs of all students. Black and Wiliam (2009) argue that diagnostic assessments are a vital component of formative assessment practices, enabling teachers to address misconceptions and support personalised learning. ASF6 (regular formative assessments) also shows significant results, highlighting the ongoing nature of assessment in supporting learning. Formative assessments provide continuous feedback and opportunities for students to improve. Popham (2008) states that formative assessment practices are integral to effective teaching as they inform both instruction and learning, helping teachers to adjust their methods and students to understand their progress.

The significance of ASF7 (engaging students in assessment discussions) reflects its role in deepening students' understanding and involvement in their learning. Engaging students in discussions about their assessments encourages them to think critically about their work and the feedback they receive. Nicol and Macfarlane-Dick (2006) suggest that such engagement promotes a more active and reflective approach to learning, enhancing students' ability to self-regulate. ASF8 (using graphic organisers) is shown to be significantly impactful in aiding students' comprehension and organisation of information. Graphic organisers help students visually structure their thoughts and understand complex concepts. According to Cañas *et al*. (2023), graphic organisers facilitate meaningful learning by helping students connect new information with prior knowledge, thus enhancing their cognitive processing. The significance of ASF9 (exit tickets) demonstrates its utility in providing quick insights into students' understanding at the end of a lesson. Exit tickets are a formative assessment tool that allows teachers to gauge the effectiveness of their instruction and identify areas that may need re-teaching. Fisher and Frey (2014) highlight that exit tickets are a simple yet powerful tool for checking students' comprehension and adjusting future lessons accordingly.

The analysis of these practices suggests that teachers' use of diverse AfL strategies significantly impacts students' learning experiences and outcomes. The implications of these findings highlight the need for professional development programmes that equip teachers with the skills to implement these practices effectively. Moreover, the inferences drawn from this analysis underscore the importance of continuous, formative assessment in creating responsive and adaptive learning environments that cater to the diverse needs of students. This aligns with the broader educational literature, which advocates for a comprehensive approach to assessment that integrates multiple strategies to support and enhance student learning.

3.1 Major Findings

1) **How do various Assessment for Learning (AfL) practices (ASF1 to ASF16) correlate with each other in terms of enhancing student learning outcomes?**

The correlation matrix demonstrates strong positive correlations between several AfL practices. For instance, ASF1 (self-assessment) shows strong positive correlations with ASF2 (peer assessment) and ASF3 (immediate feedback), with correlation coefficients around 0.80 and 0.70 respectively. This indicates that these practices are often employed together and complement each other, enhancing their overall effectiveness in improving student learning outcomes. Similar findings were noted by processes Black and Wiliam (2018), who emphasised the complementary nature of self-assessment and peer assessment in enhancing student learning.

2) **What is the impact of teacher gender on the implementation and effectiveness of AfL practices?**

The variance plot by gender shows that female teachers exhibit higher variability in implementing certain AfL practices, such as ASF1 (clear learning objectives) and ASF3 (immediate feedback). This suggests that female teachers might adopt a more context-sensitive approach, adapting these strategies based on classroom needs. Male teachers, on the other hand, show slightly higher variability in questioning techniques (ASF2), indicating a more diverse application of these methods. This observation aligns with the research by Campbell (2020), which found gender differences in the adaptability of teaching practices.

3) **How do teaching division (class teachers vs. subject teachers) and teaching experience influence the variability and effectiveness of AfL practices?**

The variance plot by teaching division reveals that class teachers exhibit higher variability in setting learning objectives (ASF1), likely due to their broader curricular responsibilities. Subject teachers show greater variability in feedback-related practices (ASF3 and ASF4), suggesting a need for more specialised training. Regarding teaching experience, more experienced teachers demonstrate lower variability across AfL practices, indicating that they have developed a consistent approach to implementing these strategies over time. This is consistent with the findings of Harris and Brown (2016), who reported that teaching experience significantly impacts the consistency of AfL practice implementation.

4) **What are the proficiency levels of teachers in utilising AfL strategies, and how do these levels vary across different teaching demographics?**

The Test Information Function (TIF) plot indicates that teachers show varied proficiency levels in utilising AfL strategies, with distinct peaks at certain proficiency levels (θ). Teachers at baseline proficiency levels ($\theta \approx 0$) are effective in foundational practices like ASF1 and ASF3, while those with higher proficiency ($\theta \approx 2$) excel in more complex strategies like ASF10 and ASF16. The gaps in the TIF curve suggest areas where additional training could improve teacher proficiency, particularly for those who fall

between or beyond the identified peaks. This mirrors the findings by DeLuca *et al*. (2019), who highlighted the importance of targeted professional development to address gaps in assessment proficiency among teachers.

3.2 Hypotheses and Outcomes

Hypothesis 1: There is a significant positive correlation between self-assessment (ASF1) and other AfL practices such as peer assessment (ASF2) and immediate feedback (ASF3).

This hypothesis is supported by the correlation matrix, which shows strong positive correlations between ASF1, ASF2, and ASF3, indicating that these practices are often used together to enhance student learning outcomes. These findings are consistent with research by Andrade and Brookhart (2020), who noted the synergistic effects of combining self-assessment with other formative practices.

Hypothesis 2: Female teachers will exhibit higher variability in the implementation of certain AfL practices compared to male teachers, indicating a more context-sensitive approach.

The variance plot by gender supports this hypothesis, showing that female teachers have higher variability in practices like ASF1 and ASF3, suggesting they adapt these strategies more to meet specific classroom needs. This is corroborated by research from Andrade *et al*. (2015), which observed that female teachers often employ more adaptable teaching strategies.

Hypothesis 3: Subject teachers will show greater variability in the use of AfL practices compared to class teachers, suggesting the need for more tailored professional development.

This hypothesis is partially supported. The variance plot shows that subject teachers have higher variability in feedback-related practices (ASF3 and ASF4), indicating a need for more specialised training to ensure consistent implementation. This aligns with findings by Brookhart and McMillan (2020), who highlighted the necessity for subject-specific professional development in assessment strategies.

Hypothesis 4: More experienced teachers will demonstrate lower variability and higher effectiveness in AfL practices, indicating that teaching experience contributes to consistent implementation.

Supported by the data, the variance plot shows that more experienced teachers exhibit lower variability across AfL practices, suggesting that experience contributes to more consistent and effective implementation. Research by Hill and Chin (2018) similarly found that teaching experience enhances the consistent application of formative assessment practices.

Hypothesis 5: Teachers with higher proficiency levels in AfL strategies, as indicated by the TIF plot, will be more effective in employing these practices, resulting in better student learning outcomes.

The TIF plot supports this hypothesis, showing that teachers with higher proficiency levels ($\theta \approx 2$) are more effective in employing complex AfL strategies like ASF10 and ASF16, leading to better student learning outcomes. The gaps in the TIF curve highlight areas for targeted professional development to enhance teacher proficiency further. These findings are supported by DeLuca *et al*. (2018), who emphasised the impact of teacher proficiency on the effectiveness of assessment practices.

The findings of this study underscore the significant role of demographic factors such as gender, teaching division, and teaching experience in the implementation and effectiveness of Assessment for Learning (AfL) strategies among basic school teachers in Ghana. The strong positive correlations between various AfL practices, such as self-assessment (ASF1), peer assessment (ASF2), and immediate feedback (ASF3), indicate that these methods are often employed in tandem to enhance student learning outcomes. These results align with the observations by Black and Wiliam (2018), who emphasised the complementary nature of self-assessment and peer assessment in fostering student engagement and understanding.

Gender differences in the implementation of AfL practices were particularly noteworthy. Female teachers exhibited higher variability in practices such as setting clear learning objectives (ASF1) and providing immediate feedback (ASF3), suggesting a more context-sensitive and adaptable approach. This finding is consistent with Graham *et al*. (2015), who found that female teachers tend to tailor their teaching strategies more to meet the unique needs of their classrooms. On the other hand, male teachers showed slightly higher variability in questioning techniques (ASF2), indicating diverse applications of these methods based on different lesson contexts or student interactions.

The analysis also revealed significant disparities between class teachers and subject teachers in their use of AfL practices. Class teachers exhibited higher variability in setting learning objectives (ASF1), likely due to their broader curricular responsibilities. In contrast, subject teachers showed greater variability in feedback-related practices (ASF3 and ASF4), highlighting the need for more specialised training in these areas. These findings echo the results of Brookhart and McMillan (2020), who reported the necessity of subject-specific professional development to ensure consistent application of assessment strategies.

Furthermore, the study highlighted the impact of teaching experience on the consistency and effectiveness of AfL practices. More experienced teachers demonstrated lower variability across AfL practices, suggesting that years of experience contribute to a more consistent and effective implementation of these strategies. This observation is supported by Harris and Brown (2016), who found that experienced teachers are better at applying formative assessment practices consistently. The use of advanced statistical techniques, such as Item Response Theory (IRT) and Test Information Function (TIF)

plots, provided nuanced insights into the proficiency levels of teachers in deploying AfL strategies. The distinct peaks and troughs in the TIF plot indicated areas of strength and gaps in teacher proficiency, suggesting targeted interventions to improve the effectiveness of AfL practices. DeLuca *et al*. (2019) similarly emphasised the importance of identifying and addressing gaps in assessment proficiency through targeted professional development.

4. Conclusion

This study provides a comprehensive analysis of how teacher demographics influence the implementation and effectiveness of Assessment for Learning (AfL) strategies in Ghana's basic schools. The findings highlight the importance of considering gender, teaching division, and teaching experience when designing and implementing professional development programmes aimed at enhancing AfL practices. Female teachers' adaptable approach and the variability seen in subject teachers' feedback practices underscore the need for tailored training that addresses specific teaching contexts and subjects. The results suggest that an integrated approach to AfL, combining self-assessment, peer assessment, and immediate feedback, can significantly enhance student learning outcomes. However, the variability observed across different demographic groups points to the necessity of continuous professional development to ensure the consistent application of these strategies. By leveraging the strengths identified through advanced statistical analyses and addressing the gaps through targeted interventions, educational policymakers and practitioners can improve the overall effectiveness of AfL practices. This study contributes to the growing body of literature on educational assessment by providing detailed insights into the interrelationships between various AfL practices and their implementation across different teaching demographics. The findings underscore the need for ongoing research and professional development to ensure that all teachers, regardless of their demographic profiles, can effectively employ AfL strategies to improve student learning outcomes. By addressing these critical factors, the educational system in Ghana can move towards more equitable and effective teaching practices that cater for the diverse needs of all students.

Acknowledgement
The author expresses his sincere appreciation to all the basic schoolteachers who attended the annual training workshop on the basic school curriculum implementation organised by the National Council for Curriculum and Assessment (NaCCA) and the Ghana Education Service (GES) in 2023.

Data Availability
The data used to support the findings of this study are available from the author upon request.

Competing Interests Statement

Author declares that no competing interests exist.

About the Author

George Oduro-Okyireh (PhD), is a Senior Lecturer in the Faculty of Education and General Studies, Akenten Appiah-Menka University of Skills Training and Entrepreneurial Development, Ghana, West Africa. He holds a PhD in Measurement and Evaluation, from the University of Cape Coast, Ghana. His research focuses on assessment instruments development with particular emphasis on the use of generalizability theory (GT) in the design of reliable assessment instruments.

References

Akhtar, M., & Saeed, M. (2020). Assessing the Effect of Agree/Disagree Circles, Exit Ticket, and Think-Pair-Share on Students' Academic Achievement at Undergraduate Level. *Bulletin of Education and Research, 42*(2), 81-96.

Andrade, H., & Du, Y. (2007). Student responses to criteria-referenced self-assessment. *Assessment & Evaluation in Higher Education, 32*(2), 159-181.

Andrade, H. L. (2010). Students as the definitive source of formative assessment: Academic self-assessment and the self-regulation of learning. In *Handbook of formative assessment* (pp. 90-105). Routledge.

Andrade, H. L., & Brookhart, S. M. (2020). Classroom assessment as the co-regulation of learning. *Assessment in education: Principles, policy & practice, 27*(4), 350-372.

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in education: Principles, policy & practice, 18*(1), 5-25.

Black, P., Harrison, C., & Lee, C. (2003). *Assessment for learning: Putting it into practice*. McGraw-Hill Education (UK).

Black, P., & Wiliam, D. (1998). *Inside the black box: Raising standards through classroom assessment*. Granada Learning.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability (formerly: Journal of personnel evaluation in education), 21*, 5-31.

Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in education: Principles, policy & practice, 25*(6), 551-575.

Bosson-Amedenu, S., Osei-Asibey, E., & Wiah, E. N. (2020). Use of assessment of learning teaching strategy among basic school teachers in Ghana. *Asian Journal of Education and Social Studies, 7*(4), 1-11.

Bromley, M. (2019). Practical classroom strategies for effective differentiation. *Seced, 2019*(11), 25-32.

Brookhart, S. M., & McMillan, J. H. (2020). *Classroom assessment and educational measurement*. Taylor & Francis.

Campbell, K. (2020). Gender and technology: Social context and intersectionality. *Handbook of research in educational communications and technology: Learning design*, 115-204.

Cañas, A. J., Reiska, P., & Shvaikovsky, O. (2023). Improving Learning and Understanding through Concept Mapping. *Knowledge Management & E-Learning, 15*(3), 369-380.

DeLuca, C., Coombs, A., & LaPointe-McEwan, D. (2019). Assessment mindset: Exploring the relationship between teacher mindset and approaches to classroom assessment. *Studies in Educational Evaluation, 61*, 159-169.

DeLuca, C., Valiquette, A., Coombs, A., LaPointe-McEwan, D., & Luhanga, U. (2018). Teachers' approaches to classroom assessment: A large-scale survey. *Assessment in education: Principles, policy & practice, 25*(4), 355-375.

DeMars, C. (2010). *Item response theory*. Oxford University Press.

El-Senousy, H. (2020). How peer assessment could be interactive and effective. *South African Journal of Education, 40*(2), 1-14.

Fisher, D., & Frey, N. (2014). *Checking for understanding: Formative assessment techniques for your classroom*. ASCD.

Graham, S., Hebert, M., & Harris, K. R. (2015). Formative assessment and writing: A meta-analysis. *The elementary school journal, 115*(4), 523-547.

Harris, L. R., & Brown, G. T. L. (2013). Opportunities and obstacles to consider when using peer-and self-assessment to improve student learning: Case studies into teachers' implementation. *Teaching and Teacher Education, 36*, 101-111.

Harris, L. R., & Brown, G. T. L. (2016). Volume introduction: The human and social experience of assessment: Valuing the person and context. In *Handbook of human and social conditions in assessment* (pp. 1-18). Routledge.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research, 77*(1), 81-112.

Heritage, M., & Wylie, E. C. (2020). *Formative assessment in the disciplines: Framing a continuum of professional learning*. Harvard Education Press.

Hill, H. C., & Chin, M. (2018). Connections between teachers' knowledge of students, instruction, and achievement outcomes. *American Educational Research Journal, 55*(5), 1076-1112.

Ketonen, L. (2021). Exploring interconnections between student peer assessment, feedback literacy and agency. *JYU dissertations*.

Lyman, F. T., Tredway, L., & Purser, M. (2023). Think-pair-share and Thinktrix: Standard bearers of student dialogue. In *Contemporary global perspectives on cooperative learning* (pp. 124-143). Routledge.

Mandouit, L., & Hattie, J. (2023). Revisiting "The Power of Feedback" from the perspective of the learner. *Learning and Instruction, 84*, 101718.

Marzano, R. J. (2012). The Two Purposes of. *Educational leadership*.

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in higher education, 31*(2), 199-218.

Novak, J. D., & Cañas, A. J. (2008). The theory underlying concept maps and how to construct and use them.

Novak, J. D., & Gowin, D. B. (1984). *Learning how to learn*. Cambridge University Press.

Osei-Asibey, E., Kusi, P., Nimoh, V., & Bosson-Amedenu, S. (2020). Evaluation of Assessment Strategies Used by Basic School Teachers in Ghana: The Case of Assessment for Learning. *Journal of Education, Society and Behavioural Science, 33*(4), 58-66.

Paulson, F. L., Paulson, P. R., & Meyer, C. A. (1991). What makes a portfolio a portfolio. *Educational leadership, 48*(5), 60-63.

Popham, W. J. (2008). *Transformative assessment*. ASCD.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional science, 18*(2), 119-144.

Stančić, M. (2021). Peer assessment as a learning and self-assessment tool: a look inside the black box. *Assessment & Evaluation in Higher Education, 46*(6), 852-864.

Stiggins, R. J. (2001). The unfulfilled promise of classroom assessment. *Educational Measurement: Issues and Practice, 20*(3), 5-15.

Stiggins, R. J., Arter, J. A., & Chappuis, J. (2004). *Classroom assessment for student learning: Doing it right, using it well*. Assessment Training Institute.

Tomlinson, C. A. (1999). Mapping a route toward differentiated instruction. *Educational leadership, 57*, 12-17.

Topali, P., Ortega-Arranz, A., Dimitriadis, Y., Villagrá-Sobrino, S., Martínez-Monés, A., & Asensio-Pérez, J. I. (2024). Unlock the Feedback Potential: Scaling Effective Teacher-Led Interventions in Massive Educational Contexts. In *Innovating Assessment and Feedback Design in Teacher Education* (pp. 1-19). Routledge.

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of educational research, 68*(3), 249-276.

Topping, K., Buchs, C., Duran, D., & Van Keer, H. (2017). *Effective peer learning: From principles to practical implementation*. Routledge.

Topping, K. J. (2010). Methodological quandaries in studying process and outcomes in peer assessment. *Learning and Instruction, 20*(4), 339-343.

Wiliam, D. (2011). *Embedded formative assessment*. Solution tree press.

Wiliam, D. (2020). Research into practice: The case of classroom formative assessment. In *Getting Evidence into Education* (pp. 119-135). Routledge.