The Assessment of Mathematical Proficiency in Written Exams: A Perspective From New South Wales (NSW)

Zehao Li
University of Sydney
zeli0824@uni.sydney.edu.au

In this study, the evaluation of mathematical proficiency (MP) in one NSW Year 12 standardised exam paper was investigated. By considering the multi-strand model of MP, a literature-based criteria for scoring the assessment of MP in individual exam questions was first developed. The criteria were then applied to quantitatively evaluate the 2023 HSC Advanced Mathematics exam paper. The findings revealed a generally balanced assessment of all MP strands in the short-response exam questions. This study has methodological implications for evaluating the assessment of MP in external mathematics assessments, and the designing mathematics exams.

In the current education landscape, there exists continuing debates around the suitability and benefits of adopting 'teach to test' pedagogies. Critiques of such pedagogies suggest that they do not enhance student's skill domains holistically in mathematics (Robinson & Dervin, 2019), and result in students to neglect enhancing their MP skills while overly focusing on obtaining strong exam marks (Sullivan, 2011). However, Seeley (2006) pointed out that if the exams in mathematics were set in a way that assesses student's MP holistically, a 'teach to test' approach should be widely adopted and accepted. Drawing from Seely (2006), educators should re-evaluate how MP is assessed in exams to better understand why certain pedagogies are preferred and respond accordingly. The tension between the desire for comprehensive MP development and 'teach to test' approaches underscores the importance of evaluating how standardised exams, such as NAPLAN and HSC exams, assess student's MP. This evaluation can shed light on the alignment between current standardised mathematics exams and contemporary pedagogical approaches, enabling educators to refine their teaching methods, assessment strategies, and potentially adjust the curriculum to satisfy more stakeholders. This study will answer the research question of whether the 2023 NSW HSC Advanced Mathematics exam paper (NESA, 2023) assesses student's MP equally. Such analysis will allow this study to illuminate the current landscape of MP assessment in exams and elucidate potential implications for future research and educational practice.

Literature Review

Mathematical Proficiency

The MP model by Kilpatrick et al. (2001) serves as a comprehensive framework capturing the fundamental attributes conducive to successful mathematics learning. This model delineates five interdependent strands: conceptual understanding, procedural fluency, adaptive reasoning, strategic competence, and productive dispositions. Enhancement of overall MP hinges upon the concurrent development of all five strands, and each strand's enhancement is contingent on the advancement of others (Schoenfeld, 2007). Since its inception, the MP model has garnered widespread acceptance and recognition as a well-established framework, as evidenced by its widespread adoption and adaption in international mathematics curricula (e.g., DOE, 2014).

In recent research the MP model has evolved into an ultimate framework that educators are expected to comprehend and integrate into their teaching practices (Cavanagh, 2021; Sullivan, 2011). Despite its widespread acceptance among teachers, researchers, and curriculum developers, ongoing debates persist regarding the efficacy of translating the model into everyday interactions within mathematics classrooms. Questions linger about the extent to

(2024). In J. Višňovská, E. Ross, & S. Getenet (Eds.), *Surfing the waves of mathematics education. Proceedings of the 46th annual conference of the Mathematics Education Research Group of Australasia* (pp. 335–342). Gold Coast: MERGA.

which educators successfully implement the model in their instructional approaches and whether it enhances student's MP in practical learning environments (Sullivan, 2011).

Mathematical Proficiency in the Australian and NSW Curriculum

In the Australian curriculum, MP is described as four strands (fluency, understanding, problem-solving, and reasoning), omitting the critical productive disposition strand described by Kilpatrick et al. (2001). The descriptions of the problem-solving and reasoning proficiencies in ACARA (2022) align with Kilpatrick's definitions of strategic competence and adaptive reasoning (Kilpatrick et al., 2001).

The NSW mathematics curriculum incorporates the content of the Australian curriculum into its own syllabuses, and its overarching key ideas align with the MP model. Departing from the terminology of MP, the NSW syllabus uses the term 'working mathematically' (NESA, n.d.). In the NSW senior mathematics syllabi, working mathematically encompasses fluency, understanding, justification, problem-solving, reasoning, and communication. The emphasis on fluency, understanding, and reasoning in both the NSW and Australian mathematics curriculum underscores their shared commitment to foundational aspects of MP (Cavanagh, 2021; NESA, n.d.). The three other working mathematically domains—justification, problem-solving, and communication—align with analogous strands in the MP model. Specifically, justification and problem-solving align with the strategic competence strand, emphasising the strategic application of existing knowledge to select appropriate and efficient methods for problem resolution (Schoenfeld, 2007). The communication skill domain is positioned as a prerequisite for every other domain, and involves using mathematical language effectively to articulate thoughts and solutions (Cavanagh, 2021). The various ways of conceptualising MP are consistent in demonstrating a movement away from pedagogy focused on rote memorisation and procedural mastery, to embracing cognitive processes such as critical thinking, reasoning, analysis, and problem-solving (Schoenfeld, 2007).

Tension Between Assessing Mathematical Proficiency and 'Teach to Test' Pedagogy

The shift in focus from rote learning to critical thinking and problem-solving in mathematics education has prompted a corresponding evolution in educators' perspectives on how and what to assess regarding student's MP (Corrêa & Haslam, 2021). Although the ideal scenario would involve the development of a mechanism capable of effectively and efficiently assessing all five strands of proficiency simultaneously, educators have encountered challenges in realising this goal. The primary challenge identified by educators is the realisation that not all MP strands lend themselves to assessment through traditional exams. Notably, the productive disposition strand, reflecting an individual's perception of mathematics as a meaningful pursuit in daily life, proves challenging to accurately assess within the confines of an exam (Corrêa & Haslam, 2021). Exams, often perceived as graduation prerequisites or progression milestones, may compel students to view mathematics as a compulsory rather than a personally rewarding activity. Consequently, assessing the productive disposition strand within exam settings becomes inherently problematic, as students may prioritise meeting societal expectations over cultivating an authentic appreciation for the subject (Sullivan, 2011).

In addition to the issue of assessing certain strands, Corrêa and Haslam (2021) highlighted the diversity of exams and their distinct purposes—ranging from diagnostic to formative, performance, and summative assessments—each serving specific roles in evaluating student learning. Different from 'well-balanced' mathematics programs, which can holistically enhance student's proficiency, mathematical thinking, and problem solving (Seeley, 2006), assessments can only assess a range of mathematical knowledge and skills (Schoenfeld, 2007). Due to stakeholders' attention on an incomplete image of student's mathematics learning revealed by formal assessments, teachers feel pressure to help students obtain satisfy results in assessments. This pressure transfers into ‘teach to test’ pedagogies, which focus solely on exam performance (Robinson & Dervin, 2019). While ‘teach to test’ pedagogies have received criticism from mathematics educators in both Australia and around the world (Sullivan, 2011; Robinson & Dervin, 2019; Seeley, 2006), this study posits that the issues created by ‘teach to test’ pedagogies are exacerbated by an unequal emphasis of MP strands in assessments. Therefore, it is necessary to evaluate how mathematics exams evaluates MP strands, and whether if there is an equal emphasis on each strand.

Methodology

The aim of this study was to determine whether exams assess all strands of MP, and measure the extent to which each strand is assessed. By doing this, this study aims to reveal whether MP strands are equally emphasised in the chosen exam paper. The 2023 NSW Advanced Mathematics exam paper is chosen for a document analysis as this subject is a prerequisite of a number of university degrees. Given that previous research had not established a reliable method to evaluate the assessment of MP in exams (Corrêa & Haslam, 2021), the first stage of this study involved the formulation of a set of criteria for determining which MP strands (and their sub-components) are assessed in an exam question. Establishing these criteria also supports reliability by providing transparency for subsequent analysis. The criteria were derived from a review of literature to establish a standardised foundation for the evaluation process (UoN, 2021; Schoenfeld, 2007). Following the development of criteria, the short response questions from the NSW Mathematics Advanced 2023 exam paper were analysed. The multiple-choice exam questions were omitted from the analysis due to the recognised limitations in assessing MP within such questions (Corrêa & Haslam, 2021). Each of the exam questions was individually analysed to determine which MP strands and their sub-components were assessed using the criteria established in the first stage of the study. Subsequently, each question was then provided with a score based on the scoring criteria, to reflect the extent to which the question assessed particular MP strand(s). Conducting this analysis will reveal whether the selected exam paper exhibits an equal emphasis on each of the MP strands in the exam. However, it is acknowledged that the evaluation of the questions is subjective in nature and open to interpretation. This subjective nature was unavoidable as there is little evidence of previous research establishing any accepted method to perform such analysis. To minimise the subjectivity when performing this evaluation in this novel situation, three trials were done to ensure the data obtained were consistent and valid.

Criteria for Evaluating the Assessment of Mathematical Proficiency in Exams

Given the limited research on the assessment of individual MP strands in standardised exams, there currently exists no rigorous framework to evaluate the assessment of mathematical proficiencies in exam questions. Therefore, establishing consistent criteria for each MP strand and assigning each question a score to reflect how they assess MP strands are necessary for this study. Thus, each exam question was analysed to determine (1) if they assess each individual proficiency strand, and (2) the depth to which the proficiency strand is assessed. To address the latter point, when scoring the depth to which each exam question assesses a particular MP strand, it will be scored at either a 1 (low level assessment of the strand), 2 (moderate assessment of the strand), or 3 (in-depth assessment of the strand). What is meant by 1, 2, and 3 specifically for each strand will be elaborated in the following subsections.

Conceptual Understanding (CU)

CU involves students developing both a robust and comprehensive knowledge of mathematics, and connections between different topics (Cavanagh, 2021). The Mathematics Assessment Resource Service (2017) illustrated that CU consists of six subcomponents: factual knowledge, comprehension, application, analysis, synthesis, and evaluation. The subcomponents have a hierarchical relationship, where factual knowledge was considered as the foundation and evaluation was considered to be the ultimate subcomponent within CU. Therefore, CU will receive a score of 1 if the question involves the application of basic factual knowledge and comprehension, a score of 2 if the question involves the application of learned knowledge in novel situations, a score of 3 if the question involves the evaluation of understanding through connecting multiple topics and synthesising new representations.

Procedural Fluency (PF)

PF captures student's ability to choose appropriate methods and carry them out efficiently, flexibly, and accurately (Schoenfeld, 2007). The three subcomponents of PF (efficiency, flexibility, and accuracy) all contribute to the overall enhancement of PF (UoN, 2021). Efficiency refers to the student's ability to select and execute a range of procedures, flexibility refers to the student's ability to select and use appropriate method(s) for solving the question, and accuracy refers to the student's ability to produce the correct answer (Schoenfeld, 2007). Therefore, PF will receive a score of 1 if the question involves the assessment of 1 identified subcomponent of PF, a score of 2 if the question involves the assessment of 2 identified subcomponents of PF, a score of 3 if the question involves the assessment of 3 identified subcomponents of PF.

Strategic Competence (SC)

SC emphasises the ability to apply knowledge in a range of situations and solve the problem accordingly (Cavanagh, 2021). SC also has three subcomponents (formulate, represent, and solve) which all contribute to the overall enhancement of SC (Schoenfeld, 2007). Formulate refers to the generation of relationships to solve the problem, represent refers to using mathematics to represent the given situation, and solve refers to evaluating results and calculating final results using appropriate methods in complex situations (Schoenfeld, 2007). Therefore, SC will receive a score of 1 if the question involves the assessment of 1 identified subcomponent of SC, a score of 2 if the question involves the assessment of 2 identified subcomponents of SC, a score of 3 if the question involves the assessment of 3 identified subcomponents of SC.

Adaptive Reasoning (AR)

AR refers to the ability to use mathematical language to communicate mathematical understanding (Cavanagh, 2021). In UoN's work (2017), AR involves three subcomponents: analysing, generalising, and justifying. Analysing refers to students exploring, comparing, and contrasting the problem and their knowledge. Generalising refers to students forming conjectures and identifying common patterns or properties. Justifying refers to students assessing the truth and making logical arguments. Therefore, AR will receive a score of 1 if the question involves assessment of 0 or 1 identified subcomponents of AR; a score of 2 if the question involves assessment of 2 identified subcomponents of AR; a score of 3 if the question involves assessment of 3 identified subcomponents of AR.

Productive Disposition (PD)

PD describes a student's ability to see mathematics as a worthwhile subject, and position themselves as capable mathematics learners (Cavanagh, 2021). However, there has been a strong emphasis on the connection between learning mathematics and daily life when discussing PD (Schoenfeld, 2007). Currently, research on how PD can be measured in school-based assessment is lacking. The development of criteria to assess PD is largely inferential and based on existing literature (Schoenfeld, 2007; Cavanagh, 2021). Therefore, the levels to which PD will be assessed in an exam question have been considered as follows: A score of 1 is given to questions that require students to solve a problem that is not related to daily life; a score of 2 is given to questions that require students to solve a problem that models a situation related to daily life; a score of 3 is given to questions that require students to solve a problem in a situation that is closely related to daily life.

Applying the Evaluative Criteria: Examples of Data Analysis Process

To provide transparency in how individual exam questions were analysed and the criteria outlined above applied, two detailed examples from 2023 NSW HSC Advanced Mathematics exam paper are discussed. It should be noted that the exam questions cannot be reproduced due to copyright restrictions, which means only descriptions of the questions can be provided. However, full questions can be accessed online (NESA, 2023).

For the first example, the question analysed is Question 15. This question assessed modelling financial mathematics concepts and involved two parts. Question 15 primarily assesses student's ability to accurately and efficiently calculate the principal and future values using the formula $A = P(1 + r)^n$. Part (a) of the question asked students to calculate the principal using the formula. Therefore, part (a) required students to apply a basic understanding of the formula, and thus scored 1 for CU. Part (a) required only a simple substitution followed by solving an equation, thus it only assessed student's accuracy in using the formula to answer the question, therefore scored 1 for PF. Part (b) of the question asked students to calculate the future value under the given condition. Therefore, part (b) required students to analyse the situation and apply their knowledge in a novel situation, thus it scored 2 for CU. Part (b) also required students to employ extra steps before substitution, thus it assessed student's efficiency and accuracy. Therefore, it was coded Level 2 for PF. Both part (a) and (b) scored for SC, as both parts assessed the ability of formulating and representing, that both items required students to generate a relationship using the formula and given information and represent such relationship using mathematical language and equations. For AR, both parts (a) and (b) scored 2 since both questions required students to analyse a given situation and generalise the situation using the formula. Despite its focus on financial mathematical concepts, this question includes some connection with student's everyday lives through the use of a problem context, thus both parts scored 3 for PD.

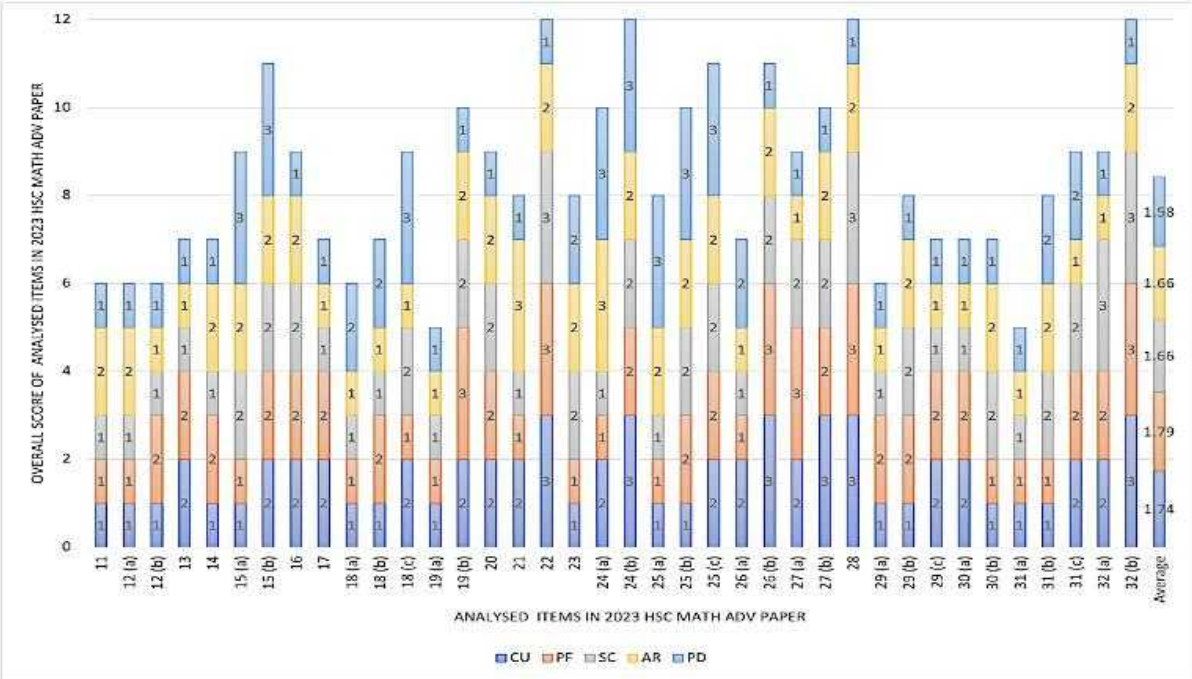
For the second example, the question analysed is Question 22. This question assessed 3D trigonometry. This item required students to utilise their knowledge of (3D) trigonometry within a complex, unfamiliar situation. Thus, it scored three for CU. This question was an excellent example of a question that assessed all three subcomponents of PF. Students were required to use different methods to perform a range of calculations accurately to get the final answer. Thus, this item scored three for PF. For SC, this question scored three. This question also tested student's ability to formulate relationships based on the given information, represent the given information in an appropriate way to obtain further information about given situation, and calculate the final answer. For AR, this item scored two since it required students to analyse given situation and generalise the situation with the use of formulas. While this question excelled in assessing various MP strands, it is noteworthy that its inherent complexity and abstract nature may limit its direct connection to real-world contexts. Thus, this item scored one for PD.

Results

The 2023 HSC Mathematics Advanced exam paper contained 39 short-response questions (including each sub-question). The analysis and final score for each question is detailed in Figure 1. Observing the overall trends in average scores, the analysis reveals the extent to which MP strands are assessed in the 2023 HSC Mathematics Advanced exam paper.

Figure 1

Coding for Each Short Response Question in the 2023 Mathematics Advanced HSC



CU obtained an average score of 1.74 across the entire exam (*SD* = 0.7), signifying that questions predominantly evaluated students' capacity to apply fundamental factual knowledge and comprehension to novel situations. PF obtained an average score of 1.79 (*SD* = 0.7), with its subcomponents—accuracy, efficiency, and flexibility—scoring 0.9, 0.5, and 0.4, respectively. This suggests a focus on accurately solving questions, while relatively fewer questions necessitate efficient and flexible problem-solving approaches. For SC, the average score was 1.66 (*SD* = 0.7). The subcomponents—formulate, represent, and solve—achieved indices of 0.5, 0.5, and 0.6, respectively, indicating a similar emphasis on assessing student's ability to formulate equations or relationships, represent situations, and solve problems within the questions. AR obtained an average score of 1.66 (*SD* = 0.6). The subcomponents—analysing, generalising, and justifying—scoring 0.6, 0.6, and 0.4, respectively, indicating a higher emphasis on assessing the ability of analysing and generalising. PD obtained an average score of 1.58 (*SD* = 0.8), denoting a relatively lower prevalence of questions testing student's application of mathematics to everyday situations. This observation highlights a potential area for consideration and further exploration in terms of incorporating real-world contexts into mathematical assessments.

Overall, the findings illustrate a similar emphasis across the five examined strands, showcasing consistent averages without significant disparities, which further reveal an equal emphasis on MP strands within the 2023 HSC Mathematics Advanced exam paper. However, as elucidated through the analysis of specific examples, such as Questions 15 and 22, and observation of the standard deviations, it becomes apparent that the exam paper incorporates a range of questions which vary in the extent to which they assess the MP strands in a balanced way. This highlights the variability in the nature of questions, ranging from those emphasising daily life connections to those with a more abstract and disciplinary focus.

Discussion

The findings from this study provide a nuanced understanding of the distribution of how MP was examined in one exam paper, shedding light on the specific areas of MP that were emphasised. Overall, the 2023 HSC Mathematics Advanced exam paper assessed MP strands in a reasonably balanced manner. However, there were several exam questions where there was a stronger emphasis on one or two strands instead of all strands.

Despite the results demonstrating a relatively balanced relationship between the five strands, PF still scored highest among all strands, particularly on its accuracy subcomponent. This finding prompts a crucial question: is this emphasis expected and/or desirable? Examining the purpose of the HSC as a means of ranking students for graduation and university entrance, the relatively higher focus on PF, especially accuracy, is an interesting and noteworthy finding. Whether this emphasis is deemed appropriate is a subject of debate. That is, this study has found that obtaining a high score in the HSC for mathematics (Advanced) and increasing one's university entrance score is largely based on the ability to accurately replicate mathematical procedures. This is potentially an oversight in the type of students that are desirable in STEM degrees, given the importance of other skills like critical thinking and problem-solving. On the other hand, an emphasis on accuracy is not entirely unexpected in a mathematics exam. What is likely desirable, however, is emphasis on accuracy (PF) as well as other strands of MP. The need for comprehensive development of MP has been argued by academics, even in the context of selective exams like the HSC (Schoenfeld, 2007). Others contend that the selective nature of the standardised exams, for instance HSC, necessitates an objective measure of student’s ability to perform tasks both accurately and effectively in the subject (Corrêa & Haslam, 2021). The purpose and focus of these HSC mathematics exams are important to consider given it influences university options in Australia.

Additionally, the findings of this study raise concerns about having an emphasis on CU or PF. While theoretically interconnected (Schoenfeld, 2007), the rise of teaching and learning practices focused on exam preparation prompts questions about whether students can enhance PF through rote learning without advancing in CU (Robinson & Dervin, 2019; Sullivan, 2011). The observed balance in the overall assessment of MP strands in this study contrasts with instances where questions assess lower CU and/or SC but higher PF. As mentioned earlier, this raises pivotal questions about the potential imbalance in the focus on proficiency strands particularly in senior years, and whether an emphasis on PF is desirable. Whether it is possible for a written format exam to more holistically assess MP is also debatable.

The study’s data prompts reflection on whether the current emphasis in exams reflects an equal attention to proficiency strands and, if not, whether adjustments are needed. Furthermore, it prompts a broader exploration of assessment mechanisms to ensure students develop critical problem-solving skills and autonomy in various situations. These questions warrant further consideration and research to inform educational practices and policies.

Conclusion

This study commenced by acknowledging the recognition among mathematics educators regarding the importance of fostering comprehensive development of student’s MP. However, as this study explored how the 2023 HSC Mathematics Advanced exam paper assesses MP strands, several noteworthy issues emerged. Despite the overall findings that there was a similar emphasis of each of the MP strands in the exam, it is noteworthy that there exists considerable difference between how different questions assess 5 MP strands. This observation prompts a critical inquiry into the suitability of questions in various types of assessment for effectively assessing MP. Ensuring that assessments align with the curriculum and pedagogical emphases advocated by policymakers, educators, and researchers is an important endeavour. By addressing these issues through rigorous research, the overarching goal of cultivating all MP strands can be more effectively realised in educational settings. The initial evaluation of one senior mathematics exam in this study highlight potential avenues for future research. It is potentially worthwhile to further explore how MP is assessed in exams at other stages of students' education. Also, it is worthwhile considering whether there are differences in how MP is assessed in school-based examinations compared to externally set exams such as the HSC. Such research could provide insights into whether the current emphasis on enhancing MP is identical across various educational levels and assessment types.

The assertion made by Schoenfeld (2007) regarding the varied interpretations of the MP model among educators holds significant relevance for this study. This contention becomes particularly pertinent in the context of this study, where the evaluation of questions to assess MP strands is a central focus. While there is widespread consensus on the importance of enhancing MP and its integration into teaching practices (Corrêa & Haslam, 2021; Sullivan, 2011), the absence of a universally accepted method for evaluating how MP strands are assessed in exam papers is a notable gap in the existing literature. In response to this gap, the current study has contributed to methodologies in this area by developing a set of criteria for analysis that is grounded in the existing understanding of MP strands. However, the acknowledgment that other researchers may hold different opinions regarding the details of the evaluation mechanism may influence the validity and reliability of this study. This acknowledgment underscores the need for continued dialogue and refinement in methodologies for evaluating MP in exam papers, with the aim of establishing more standardised and widely accepted assessment frameworks in the future.

References

Australian Curriculum, Assessment and Reporting Authority (ACARA). (2022). *Australian Curriculum: F–10 Mathematics Curriculum (v9.0).* ACARA. https://v9.australiancurriculum.edu.au/

Cavanagh, M. (2021). Introduction: the learning and teaching of mathematics. In G. Hine, R. Reaburn, J. Anderson, L. Galligan, C. Carmichael, M. Cavanagh, B. Ngu, & B. White (Eds.), *Teaching secondary mathematics* (pp. 8–34). Cambridge University Press.

Corrêa, P., & Haslam, D. (2021). Mathematical proficiency as the basis for assessment: A literature review and its potentialities. *Mathematics Teaching Research Journal*, *12*(4), 3–20.

Department of Education (DOE). (2014). Mathematics programmes of study: Key stage 4 National curriculum in England. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/331882/KS4_maths_PoS_FINAL_170714.pdf

Kilpatrick, J., Swafford, J., & Findell, B. (Eds.). (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: The National Academies Press.

Mathematics Assessment Resource Service. (2017). Developing mathematical proficiency: The potential of different types of tasks for student learning. [Conference presentation]. University of Nottingham. https://www.mathnic.org/downloads/06_proficiency_slides.pdf

NSW Education Standards Authority [NESA]. (2023). *Mathematics advanced 2023 HSC exam pack*. NESA. https://www.educationstandards.nsw.edu.au/wps/portal/nesa/resource-finder/hsc-exam-papers/2023/mathematics-advanced-2023-hsc-exam-pack

NSW Education Standards Authority [NESA]. (n.d.). *Working mathematically*. https://educationstandards.nsw.edu.au/wps/portal/nesa/11-12/stage-6-learning-areas/stage-6-mathematics/mathematics-standard-2017/working-mathematically

Robinson, B. K., & Dervin, A. (2019). Teach to the student, not the test. *Urban Education Research & Policy Annuals*, *6*(2).

Schoenfeld, A. (2007). What is mathematical proficiency and how can it be assessed? In A. Schoenfeld (Ed.), *Assessing mathematical proficiency* (pp. 59–74). MSRI Publications & Cambridge University Press. https://doi.org/10.1017/CBO9780511755378

Seeley, C. (2006). *Teaching to the test*. National Council of Teachers of Mathematics. https://www.nctm.org/News-and-Calendar/Messages-from-the-President/Archive/Cathy-Seeley/Teaching-to-the-Test/

Sullivan, P. (2011). *Teaching mathematics: Using research-informed strategies*. Australian Council for Educational Research.