The Mathematics Educator
2019, Vol.18, No.1&2, 1-21

Mathematics Teachers' Perceptions of the Cognitive Demand of Assessment Items

NG Wee Leng, TAN Hwee Chiat June, Paul M. E. SHUTLER

National Institute of Education
Nanyang Technological University
Singapore

Abstract: Teachers' judgements of the cognitive demand of mathematical assessment items have implications for the nature of students' learning experiences. The study examines Advanced-level (or Grades 11 and 12) teachers' perceptions of the cognitive demand of mathematical assessment items and explores the relationships between teachers' perceptions of cognitive demand and students' learning. Fourteen A-level mathematics teachers and 66 Grade 11 students from two schools in Singapore participated in this study. Findings from this study showed that when judging the cognitive demand of items, teachers considered the complexity level of an item, the inherent difficulty of the concepts being tested, and students' familiarity with the item. This study also found that teachers' perceptions of the cognitive demand of items differed due to their gender, teaching experience, and academic qualifications.

Keywords: Cognitive demand, assessment, Advanced-level mathematics teachers

Introduction

Over the past several decades, research on mathematical thinking and learning has re-conceptualised what mathematics learning and mathematical competency should be (e.g., Gardner, 1985; Schoenfeld, 1985). As a result, mathematics researchers and educators have called for a shift in emphasis in teaching of mathematics to developing mathematical processes such as mathematical reasoning and communication rather than computational skills alone. For instance, the Common Core State Standards for Mathematics highlight the ability to reason abstractly and quantitatively and the ability to construct viable arguments and critique the reasoning of others as mathematical practices that teachers should emphasise in mathematics classrooms (National Governors Association Center for Best Practices, Council of Chief State School Officers, 2010). To help students effectively develop these mathematical reasoning processes, the type of activities and tasks that they work on would need to change. For example, the tasks that students work on in class should encourage them to think and make sense of mathematics in meaningful ways, and tasks that students tackle in assessment items should assess mathematical reasoning skills and not routine procedures and memory work alone.

Gorin and Embretson (2013) emphasised the need for teachers to be more aware of the cognitive processes that students use for answering assessment items. With a greater awareness of the cognitive demand of a task, teachers could develop a more refined mental model of how students think and reason mathematically, enabling them to gain insights into possible sources of difficulties that students are likely to encounter when solving problems. Schoenfeld (1988) found that the mathematical tasks that teachers used were predominantly those that were closely linked to or modelled after the types of questions that students would encounter in standardised tests or examinations. Thus, it is important that mathematics teachers are able to identify and classify the assessment items that they use for classroom activities in terms of how cognitively demanding the items are. In this way, they can ensure that the tasks that they use for class activities are appropriately challenging in terms of the cognitive processes in order to support students' development of mathematical processes.

However, much of the knowledge that teachers use in teaching is tacit (Verloop, Van Driel, & Meijer, 2001), and little is known about teachers' perceptions of what counts as cognitively demanding in mathematics. Intuitively, one would think that teachers with more years of teaching experience could be more adept at anticipating whether students are likely to find an assessment item easy or difficult, be more familiar with how students are likely to approach solving the items, and where students are likely to go wrong. Thus, it is possible that teachers' perceptions of the cognitive demand of mathematical tasks could be influenced by their background characteristics. For example, Blömeke and Delaney (2012) found that male teachers were less attuned than female teachers to how students think and the misconceptions that students may have. This suggests that male and female teachers could have different perceptions of the cognitive demand of mathematical tasks that they use.

Researchers in the area of teaching and teacher education have often compared novice or preservice teachers with their more experienced colleagues. For example, using academic qualifications as an indirect measure of teachers' professional knowledge, Goldhaber and Brewer (2000) found that teachers' academic qualifications are positively correlated with student achievement at the secondary level. Rowan, Chiang, and Miller (1997) measured mathematics teachers' content knowledge (CK) in terms of their high school mathematics knowledge and found a positive association between teacher's CK and students' achievement. The importance of teachers' CK was also highlighted by Ball (1990), who found that teachers' ability to represent and explain content to students is limited by their CK. Ball also found that preservice mathematics teachers do not have adequate CK to teach for mathematical understanding. Taken together, these studies indirectly suggest that teachers with different academic qualifications could also perceive the cognitive demand of mathematical tasks or items differently possibly because of their CK, since CK is needed for understanding and unpacking the cognitive demand inherent in the task.

According to Pollitt, Ahmed, and Crisp (2007), examiners and subject matter experts draw on their professional knowledge when judging the cognitive demand of assessment items. In the assessment context, the cognitive demand of an item refers to the level of thinking that a student is assumed to have and be able to perform in order to accomplish the task (Pollitt et al., 2007). Likewise, when gauging the cognitive demand of an assessment item and selecting it for students to work on, teachers need to understand the cognitive processes expected of students in order to solve the task set out in the assessment item. They must also consider the instructional and assessment purposes the item is meant to serve when it comes to students' development of mathematical understanding, processes, and skills. According to Ball, Lubienski, and Mewborn (2001), it is teachers' PCK that underlies the tasks that teachers do such as, constructing and choosing tasks, and monitoring students' understanding. From this perspective, teachers would need to draw on specific aspects of their professional knowledge such as, their subject matter knowledge (SMK), knowledge of their own students, and their pedagogical knowledge (Ball, Thames, & Phelps, 2008), when gauging the cognitive demand of mathematical assessment items. In addition, several researchers have found that teaching experience contributes to teachers' PCK (e.g., Grossman, 1990; Van Driel & Berry, 2009). In fact, Van Driel and Berry identified preservice teachers' lack of teaching experience as the reason for preservice teachers' inferior PCK and suggested that until preservice teachers gained experience and confidence in teaching, their development of PCK would be delayed.

It would appear that the process of judging the cognitive demand of assessment items could be influenced by teachers' individual characteristics (e.g., their teaching experience and qualifications), and to some extent, their professional knowledge. With the emphasis on developing students' mathematical reasoning, and that teachers tend to use past assessment items to guide their instructional practices, it is important that they are able to identify and determine the level of cognitive demand in the assessment items that they use.

In the context of Singapore, Advanced-level (or Grades 11 and 12) mathematics teachers tend to take reference from past Advanced-level (A-level) examinations in preparing their instructional materials. There is no standardised textbook for A-level Mathematics and teachers often use past examination items as practice problems for students to work on. How do A-level mathematics teachers gauge the cognitive demand of the assessment items that they use? In this study, we will examine teachers' perceptions of the cognitive demand of A-level mathematical assessment items and investigate the effects of teachers' characteristics on their perceptions. More specifically, we address the following questions: What factors do A-level mathematics teachers consider when judging the cognitive demand of A-level mathematical assessment items? Is there a significant difference in perceptions of cognitive demand among teachers with different background characteristics (i.e., years of teaching experience, academic qualifications, and gender)? What are the relationships amongst different dimensions of cognitive demand and students' understanding of mathematics?

Method

Participants

A convenience sample of 14 teachers (nine females and five males) and 66 Grade 11 students from two schools in Singapore that offer the A-level curriculum was used in this study. There were nine teachers and 25 students from School A, and from School B, there were five teachers and 41 students. The number of years of teaching experience among the teachers ranged from one to 34. A majority of the teachers (85.7%) had at least a bachelor degree with honours. The students involved in this study were from five different mixed-ability classes that were taught by five of the teachers involved in the study.

Instruments and Measures

Cognitive demand instrument. Teachers' perceptions of the cognitive demand of assessment items were measured at two levels: (1) overall demand and (2) specific dimensions of cognitive demand. Overall demand was assessed at three levels: low, moderate, and high. In addition, teachers were asked to provide the reasons underlying their ratings of overall demand. The specific dimensions of cognitive demand were: complexity, abstractness, and strategy. Teachers' perceptions of cognitive demand in these three dimensions were measured using a cognitive demand instrument developed by the Tan, Ng and Shutler (2017) specifically for gauging the cognitive demand of A-level mathematical assessment items. The cognitive demand instrument was adapted from the Complexity-Resources-Abstractness-Strategy framework by Hughes, Pollitt, and Ahmed (1998).

The cognitive demand instrument used in this study has five levels of demand for each of the three dimensions. The first dimension, complexity, refers to the demand placed on students in terms of the sub-goals or set of steps, specified (or implied) in a question to guide them. The second dimension, abstractness, is the demand placed on students in terms of working with abstract elements, which are mathematical objects and ideas that require considerable mental construction or some imagination (e.g., variable coefficients, unknown functions, asymptotes and limits), rather than concrete elements, which are completely specified (e.g., numerical coefficients, visible intersections). The third dimension, strategy, refers to the demand placed on students in terms of how they decide to tackle a question.

Teacher questionnaire. Teachers' perceptions of cognitive demand were also gathered using a questionnaire so that we could understand the reasons as to how teachers go about gauging the demand of the items. The questionnaire comprises two sections: (a) questions related to basic demographic information (e.g., gender and number of years of teaching experience), and (b) questions that were intended to elicit teachers' views on mathematical assessment items (e.g., how teachers usually judge if a question is easy or challenging).

Mathematical assessment items. The mathematical assessment items used in this study were selected from the 1999 to 2015 GCE A-level Mathematics examinations for Singapore candidates. A total of 12 items that assess concepts and skills from various Pure Mathematics topics (e.g., Function, Inequalities, and Calculus) were used.

Procedure

All the teachers were asked to complete the questionnaire and given about one week to rate the overall demand of the set of 12 mathematical assessment items. Teachers were asked to work independently and refrain from discussing their ratings with their colleagues who were also involved in the study. Following this, the teachers attended a training session to familiarise them with the cognitive demand instrument. Similarly, they were asked to work independently and given one week to rate the same set of 12 items but this time, in terms of the three cognitive demand dimensions. The students were given a set of items as a take-home assignment. As the students from School A had learnt three topics, namely Functions and Graphs, Equations and Inequalities, and Sequences and Series, they were given seven items on these three topics to solve. The seven items were a subset of the set of items given to the teachers to rate. School B students had covered only two topics, namely Functions and Graphs, and Equations and Inequalities. Thus, the students from School B were given only the five items on these two topics to solve, these being the same items that were given to students from School A. The students' solutions were collected after about a week.

Results

Reasons Underlying Teachers' Judgement of Cognitive Demand

Teachers judgements of cognitive demand were influenced by their perceptions of the complexity level of an item and the process skills that were required for solving the task set up in the item. For instance, items that involved direct application of a formula, simple recall or numerical calculations were considered by the participating teachers to be lower in demand. On the other hand, higher demand items were those that required "deeper analysis", "critical thinking or some logical reasoning", or "extension beyond basic concepts". Additionally, the teachers thought that assessment items that required application of more than one topic would increase the complexity level of the item. Besides the complexity level of an item, the inherent difficulty of concepts in mathematics was another consideration that seemed to influence teachers' judgements of cognitive demand. For example, concepts of modulus functions, limits and integration were viewed by teachers as abstract for A-level mathematics students. As such, items that assessed these concepts were judged to be of higher demand by the teachers. In addition, teachers also indicated that the use of unknowns and having to deal with mathematical symbols and notations could raise the demand level of an item for students. Another factor that teachers used in their judgements of cognitive demand was whether or not students had "seen" or had prior exposure to the item in the course of learning mathematics. They described an "unseen" item as a question that students had not encountered before, whereas a "seen" item was one that students would most likely find familiar because the students had encountered a task or question of similar nature before. Compared to an "unseen" assessment item, a "seen" item was considered as being lower in cognitive demand.

Teacher Characteristics and their Perceptions of Cognitive Demand

Academic qualifications and cognitive demand perceptions. A series of Friedman tests with randomised block design were conducted to evaluate the effect of academic qualification level on teachers' cognitive demand ratings. The independent variable, academic qualification, had three levels: bachelor, bachelor with honours, and master's degree. The means and standard deviations for the ratings by the participating teachers with different academic qualification levels are shown in Table 1.

Table 1
Means and Standard Deviations of Teachers' Ratings by Academic Qualifications

Cognitive Demand	Bachelor (n = 2)		Bachelor (Hons) (n = 9)		Master's (n = 3)	
	M	*SD*	*M*	*SD*	*M*	*SD*
Complexity	1.42	0.63	2.38	0.77	2.21	0.69
Abstractness	1.50	0.71	2.17	0.72	2.33	0.81
Strategy	1.71	0.86	2.21	0.66	2.08	0.87
Overall	2.13	0.61	1.96	0.81	1.92	0.73

The results indicated that there were no significant differences in ratings for overall demand among the three academic qualification levels, $\chi^2(2) = .400$, $p = .819$. Although the differences were not significant, participating teachers with a bachelor degree rated the overall demand of the items highest, whereas those with a master's degree perceived the overall demand of the items to be lowest. Similarly, no significant differences were found in the ratings among teachers with different levels of academic qualifications for abstractness, $\chi^2(2) = 4.87$, $p = .088$, and for strategy, $\chi^2(2) = 1.70$, $p = .428$. In contrast to the ratings for overall demand, participating teachers with a bachelor degree rated the demand in abstractness lowest, whereas those with a master's degree gave the highest ratings for this dimension. In other words, the general trend was reversed, although there were no statistically significant differences among teachers with different qualification levels.

Teachers with different academic qualification levels differed significantly in their ratings of complexity demand, $\chi^2(2) = 6.71$, $p = .035$. Those with honours degrees perceived the demand in complexity to be highest ($M = 2.38$, $SD = 0.77$), followed next by teachers with a master's degree ($M = 2.21$, $SD = 0.69$), and then those with a bachelor degree ($M = 1.42$, $SD = 0.63$). Post-hoc Dunn-Bonferroni tests (using alpha level of .017) were conducted to evaluate the three pairwise differences in complexity ratings. However, no significant differences were found between teachers with a bachelor degree and those with honours ($p = .019$), between those with bachelor degree and master's degree ($p = .052$), as well as between those with honours and master's degree holders ($p = .683$).

Years of teaching experience and cognitive demand perceptions. Friedman tests with randomised block design were also conducted to evaluate the effect of the number of years of teaching experience on teachers' cognitive demand ratings. The independent variable, years of teaching experience, had three levels: 1 to 10 years, 11 to 19 years, and at least 20 years. This classification of years of teaching experience ensured that there was an approximately equal number of teachers in each group or level. In the Singapore context, teachers who had taught for more than 10 years were generally considered as being more senior or experienced. The means and standard deviations for the ratings by the teachers with different years of teaching experience are shown in Table 2.

Table 2
Means and Standard Deviations of Teachers' Ratings by Years of Teaching Experience

Cognitive Demand	1 to 10 Years (n = 4)		11 to 19 Years (n = 6)		At least 20 Years (n = 4)	
	M	*SD*	*M*	*SD*	*M*	*SD*
Complexity	2.38	0.77	1.50	0.48	2.13	0.88
Abstractness	2.04	0.69	1.54	0.78	2.42	0.67
Strategy	2.04	0.62	1.21	0.33	2.75	0.40
Overall	2.17	0.83	2.00	0.48	1.83	0.86

The tests showed that the number of years of teaching experience did not have a significant effect on the teachers' ratings for complexity, $\chi^2(2) = 5.20$, $p = .074$, for abstractness, $\chi^2(2) = 5.16$, $p = .076$, and for overall demand, $\chi^2(2) = 0.84$, $p = .656$. For complexity and overall demand, the highest mean ratings were from teachers with 1 to 10 years of teaching experience. In other words, compared to those with more years of teaching experience, teachers with less experience tended to perceive the assessment items to be more demanding in the complexity dimension and at the overall demand level. In contrast, the lowest ratings for complexity and overall demand were from participating teachers with 11 to 19 years of experience and those with at least 20 years of experience, respectively. For abstractness, the highest ratings were from the teachers with at least 20 years of teaching experience, while the lowest ratings were from the group with 11 to 19 years of teaching experience.

For strategy, there were significant differences in ratings among teachers with different number of years of teaching experience, $\chi^2(2) = 15.95$, $p < .001$. Post-hoc Dunn-Bonferroni tests (using alpha level of .017) were conducted to evaluate the three pairwise differences in ratings for strategy. The teachers with 11 to 19 years of teaching experience rated the items less demanding for strategy compared to those with at least 20 years of experience ($p < .001$). In other words, teachers in the group with the most number of years of teaching experience were of the opinion that the assessment items were more demanding in the strategy dimension in comparison to those with between 11 to 19 years of teaching experience. However, teachers with 1 to 10 years of experience did not differ significantly in their ratings of demand compared to those with 11 to 19 years of experience ($p = .041$), and also with those having at least 20 years of experience ($p = .083$).

Treating years of experience as interval data, ordinal regression analyses were used to model the relationships between years of teaching experience with each of the cognitive demand dimensions as well as overall demand. In Model 1, with complexity as the outcome variable, the test of parallel lines was not significant, $\chi^2(3) = 4.30$, $p = .231$. The overall model fit was not significant, $\chi^2(1) = 4.50$, $p = .480$, but the deviance chi-square statistic was significant ($p = .038$). Moreover, the Nagelkerke R^2 was very small, with only 0.3% of the variance in complexity accounted for by years of teaching experience. These results indicate that teachers' years of teaching experience do not significantly explain their ratings of complexity demand. Similarly, the test of parallel lines was nonsignificant in Model 2, where abstractness was the outcome variable, $\chi^2(3) = .86$, $p = .835$. The overall model fit was not significant, $\chi^2(1) = 1.06$, $p = .304$, and the deviance chi-square statistic was not significant ($p = .298$). The Nagelkerke R^2 was also very small, indicating that only 0.7% of the variance in abstractness was accounted for by years of teaching experience. These results suggest that years of teaching experience is not a significant predictor of teachers' perceptions of demand in the abstractness dimension.

Model 3 was used to explore the effect of teachers' years of teaching experience on their strategy ratings. The test of parallel lines was not significant, $\chi^2(3) = 1.71$, $p = .635$. The overall model was not a good fit, $\chi^2(1) = .862$, $p = .353$, but the deviance chi-square statistic was significant ($p = .039$). The Nagelkerke R^2 was also very small, indicating that only 0.5% of the variance in strategy was explained by years of teaching experience. Thus, teachers' years of teaching experience is also not a significant predictor of their perceptions of demand in strategy. In Model 4, the outcome variable was overall demand. The test of parallel lines was not significant, $\chi^2(3) = .41$, $p = .521$. Likewise, the model was not a good fit to the data, $\chi^2(1) = .00$, $p = .984$, and the deviance chi-square statistic was not significant ($p = .342$). The Nagelkerke R^2 was almost negligible, which indicate that teachers' years of teaching experience did not explain any of the variance in overall demand. Thus, teachers' years of teaching experience does not have a significant effect on their perceptions of overall demand. In summary, all the four models indicated that teachers' years of teaching experience do not significantly explain their ratings in complexity, abstractness, strategy and overall demand. The model fit values and parameter estimates for variables are summarised in Table 3.

Table 3
Ordinal Regression Model Fit and Parameter Estimates

Variable	Model Fit	*B*	Wald χ^2-test	Odds Ratio	95% CI
Model 1	137.59			1.01	
Complexity L1		–1.22	17.86***		[–1.78, –0.65]
Complexity L2		–0.30	1.28		[–0.83, 0.22]
Complexity L3		1.16	16.82***		[0.61, 1.72]
Complexity L4		2.73	52.62***		[1.99, 3.47]
Teaching experience (in years)		0.01	0.55		[–0.02, 0.04]
Model 2	124.62			1.02	
Abstractness L1		–1.43	22.75***		[–2.01, –0.84]
Abstractness L2		0.28	1.09		[–0.25, 0.81]
Abstractness L3		1.85	35.35***		[1.24, 2.46]
Abstractness L4		3.24	56.21***		[2.39, 4.08]
Teaching experience (in years)		0.02	1.11		[–0.01, 0.05]
Model 3	135.55			1.02	
Strategy L1		–1.24	18.34***		[–1.81, –0.67]
Strategy L2		0.19	0.49		[–0.34, 0.72]
Strategy L3		1.71	31.65***		[1.11, 2.30]
Strategy L4		3.52	54.44***		[2.59, 4.46]
Teaching experience (in years)		0.02	0.93		[–0.02, 0.04]
Model 4	68.33			1.00	
Overall Demand L1		–1.27	17.72***		[–1.87, –0.68]
Overall Demand L2		0.76	6.85		[0.19, 1.32]
Teaching experience (in years)		0.00	0.00		[–0.03, 0.03]

Note. L1 = Level 1; L2 = Level 2; L3 = Level 3; L4 = Level 4. The model fit values refer to the –2 Log-likelihood values for the final model with years of teaching experience as the predictor. *** $p < .001$

Gender and cognitive demand perceptions. To evaluate the effect of gender on teachers' cognitive demand ratings, a series of Friedman tests with randomised block design were conducted. The means and standard deviations of the ratings of the items are shown in Table 4. The tests indicated that there were no significant differences in ratings between male and female teachers for overall demand, $\chi^2(2) = 1.600$, $p = .206$. However, male and female teachers' ratings were significantly different for complexity, $\chi^2(2) = 8.33$, $p = .004$, abstractness, $\chi^2(2) = 5.33$, $p = .021$, and strategy, $\chi^2(2) = 4.46$, $p = .035$. Compared to male teachers, female teachers rated the demand of the items higher for all three of these dimensions. In other words, female teachers tended to perceive the assessment items to be more demanding in the complexity, abstractness, and strategy dimensions than the male teachers.

Table 4
Means and Standard Deviations of Teachers' Ratings by Gender

	Males (n = 5)		Females (n = 9)	
Cognitive Demand	*M*	*SD*	*M*	*SD*
Complexity	1.08	0.29	1.92	0.29
Abstractness	1.17	0.39	1.83	0.39
Strategy	1.21	0.40	1.79	0.40
Overall	1.33	0.44	1.67	0.44

Note. Ratings of overall demand were provided by only eight female participating teachers as one of the female teachers did not provide her ratings.

Relationships Among Cognitive Demand Dimensions and Students' Understanding of Mathematics

Spearman rank-order correlations among the 14 teachers' ratings of cognitive demand and mean scores of the 66 students for the seven items are shown in Table 5. In particular, the correlation between scores and abstractness was the strongest, while the correlation between scores and complexity was the weakest. Regression analyses were conducted to investigate the relationships between teachers' perceived level of demand, both for the three dimensions and also overall demand, and students' scores on the items. The analyses were conducted in two distinct steps: firstly, by considering the median ratings of all the teachers and the mean scores of all participating students; secondly, by narrowing the focus to only the teachers who taught the five classes of students.

Table 5
Inter-correlations among Cognitive Demand Dimensions and Student Scores

Cognitive Demand	1	2	3	4	5
1. Complexity	—	—	—	—	—
2. Abstractness	.464	—	—	—	—
3. Strategy	.847*	.847*	—	—	—
4. Overall	.955**	.595	.891**	—	—
5. Scores	–.286	–.714	–.559	–.378	—

Note. **p* < .05. ** *p* < .01.

Median cognitive demand ratings and mean student scores. For each of the seven items, mean student scores were calculated and converted to percentages (out of the total number of marks per item). The median ratings given by the 14 teachers were computed for the overall demand and for each of the dimensions for each item. Categorical regression analyses were conducted to predict mean student scores from each of the three dimensions of demand and from overall demand. Both complexity ($\beta = -.791$, $p = .225$) and strategy ($\beta = -.791$, $p = .142$) were not significant predictors of student scores, whereas abstractness ($\beta = -.681$, $p = .050$) and overall demand were significant predictors of student scores ($\beta = -.791$, $p = .001$).

Cognitive demand ratings of teachers and their classes. As the students from School A were from two classes taught by the same teacher, for the purpose of data analysis, the students' scores were analysed as one class. Mean scores for each of the five classes of students were computed and converted to percentages (see Table 6). Instead of using the entire sample of 14 teachers, only ratings from the five teachers who taught the 66 students were used in the regression analyses. In addition, only the five items that were attempted by all 66 students were considered in the analyses. This ensured that the total number of items and marks were the same for each class.

Table 6
Mean Student Scores by Class

	School A	School B			
Item	Class 1 (n = 25)	Class 1 (n = 5)	Class 2 (n = 11)	Class 3 (n = 12)	Class 4 (n = 13)
1	96.00	60.00	86.35	87.50	92.30
2	88.67	100.00	96.97	84.72	87.18
3	58.00	62.50	60.23	64.59	60.58
4	65.20	80.00	51.82	69.58	74.23
5	82.67	33.33	13.63	54.17	60.25

To examine the relationships between students' scores and the ratings of demand given by their teachers, multiple regression analyses were conducted. Regressing scores on complexity, the results were significant, $F(3, 21) = 3.478$, $p = .034$. The three complexity variables explained 33.2% of the variance in student scores. Follow-up post hoc comparisons using Dunnett's test (with $t(20) = 2.54$ at $\alpha = .05$) showed that for complexity, the differences in demand between Level 1 and Level 2 ($t = 3.85$) as well as between Level 1 and Level 3 ($t = 2.62$) had significant influence on student scores. There were no statistically significant differences between the other four pairs of comparisons across the levels of complexity demand.

Similarly, regressing scores on strategy, the results were significant, $F(3, 21) = 3.568$, $p = .031$. The three strategy variables accounted for 33.8% of the total variance in student scores. Follow-up post hoc comparisons using Dunnett's test (with $t(20) = 2.54$ at $\alpha = .05$) showed that for strategy, only the differences in demand between Level 1 and Level 4 ($t = 2.754$) had a significant influence on student scores. Differences between the other five pairs of comparisons across the levels of strategy demand were not statistically significant. Furthermore, both abstractness, $F(2, 22) = 2.000$, $R^2 = .154$, $p = .159$, and overall demand, $F(2, 22) = 1.590$, $R^2 = .126$, $p = .226$, did not have any significant influence on student scores. In addition, categorical regression analyses were conducted to examine the relationships between mean student scores and each of the three dimensions of demand and for overall demand. Complexity ($\beta = -.576$, $p < .001$), strategy ($\beta = -.546$, $p < .001$), and overall demand ($\beta = -.356$, $p = .040$) were significant predictors of student scores, while abstractness ($\beta = -.411$, $p = .105$) was not a significant predictor.

Discussion

The findings from this study provide insights into how teachers perceive cognitive demand in mathematics and the tacit knowledge that guide their professional judgements of cognitive demand. Teachers' judgements of cognitive demand were influenced by their perceptions of the complexity of processes and subject or concept difficulty. Essentially, these two factors, as described by the participating teachers, can be classified under the dimensions of complexity, abstractness, and strategy as described in the cognitive demand instrument. Although the participating teachers thought that a seen versus unseen problem made a difference in their perception of cognitive demand, this pertained more towards the difficulty level of a task or an item rather than its inherent cognitive demand. The teachers could have thought that when students encountered a task or item that they had seen before, they would have a mental model of how the item could be answered (Gorin & Embretson, 2013), and this could then make the item easier or less challenging.

The teachers' mean ratings of the overall demand and the three cognitive demand dimensions for the items used were significantly correlated, suggesting that if they perceived the overall demand of an item to be high, it was likely that they would perceive the demand in the complexity, abstractness, and strategy dimensions of the item to be high as well. Likewise, the significantly high correlations among the complexity, abstractness, and strategy dimensions suggest that the teachers tended to perceive the demands in these three dimensions as interrelated and were likely to associate a high (or low) demand in one dimension with also high (or low) demand in the other two dimensions. A possible explanation is that teachers were unable to successfully distinguish the demand across the three dimensions, even though they attended a training session on how to use the instrument to judge the cognitive demand of A-level mathematical assessment items. It could also be the case that the teachers' judgements of cognitive demand of the assessment items were affected by their perceptions of how their students would tackle and solve the items when in fact, the teachers should focus solely on the cognitive demand of the items as intended by the examiners. Another possible explanation is that given the purpose and constraints of national examinations, there might not be wide variations (resulting in a lack of correlation) in demand across the three dimensions.

Overall, it appeared that the teachers have a broad idea of what constituted cognitive demand in A-level mathematics. However, they did not necessarily differentiate between cognitive demand and difficulty level. This is not surprising as teachers tend to be more concerned about whether their students can obtain the correct answer to a question and score well in tests and examinations. It could be the case that the teachers had perceived the cognitive demand of an item in terms of its effect on the difficulty level and on how students would actually attempt to answer the items, rather than on the cognitive demand intended by the examiner per se.

Effects of Teachers' Background Characteristics

This study also investigated if teachers' perceptions of cognitive demand were influenced by their background characteristics. There was significant gender effect on teachers' perceptions of the cognitive demand of mathematical assessment items in the complexity, abstractness, and strategy dimensions. Compared to the male participating teachers, female teachers were more likely to consider the demand in these dimensions as higher. This could be because female teachers are more likely to be attuned to how students think and the errors that students are likely to make as pointed out by Blömeke and Delaney (2012), which in turn, could have influenced their judgements of cognitive demand. Moreover, as teachers' beliefs and perceptions of their students' abilities influence their teaching (Porter & Brophy, 1988), it is possible that the teachers' perceptions of their students could have affected their decisions about the cognitive demand of mathematical assessment items.

Teachers' qualifications also had a significant effect on their perceptions of the complexity of the assessment items. Teachers with an honours degree were most likely to rate the demand in complexity as higher, compared to those with a master's degree or a bachelor degree. Teachers' qualifications, however, did not seem to significantly affect their ratings of the other cognitive demand dimensions. Years of teaching experience had a significant effect on teachers' perceptions of the demand in the strategy dimension. The most experienced group of teachers, with at least 20 years of service, were more likely to rate the demand level in the dimension of strategy higher than teachers with 11 to 19 years of experience. However, years of teaching experience was not a significant predictor of teachers' ratings of overall demand, nor of the three dimensions separately.

Taking these findings together, it would seem that teachers perceive the cognitive demand of items differently due to their differing teaching experience or academic qualifications. Even though all the teachers indicated that they relied on their past teaching experiences in judging the cognitive demand of almost every item, the nature of their teaching experiences would most likely be different. Teachers with fewer years of experience might perceive the items to be less demanding cognitively because they are unfamiliar with the curriculum or they do not have indepth knowledge of their students' abilities. In contrast, teachers with more years of teaching experience are likely to have richer PCK (Grossman, 1990; Van Driel & Berry, 2009) as well as more extensive knowledge of the curriculum content and students' reasoning, errors, and learning styles (Ball et al., 2008). The teachers also indicated that they relied on their CK (or SMK) to gauge the cognitive demand of the assessment items. Given that teachers' ability to explain the content to students is limited by their CK (Ball, 1990), it could be the case that teachers with different qualifications would perceive the cognitive demand of items differently because of their CK, which in turn, could have an effect on their ratings.

Links between Cognitive Demand Perceptions and Students' Learning

Given that the cognitive demand inherent in assessment tasks have an impact on the difficulty of the tasks for students (Crisp & Novaković, 2009; Pollitt et al., 2007), this study also explored the relationship between teachers' perceptions of cognitive demand and students' understanding of mathematics. Correlations between teachers' mean perceptions of cognitive demand and students' mean scores were negative and ranged from weak to strong, although they were not statistically significant. These results suggest that higher demand could lead to more difficulty, which in turn, translates to lower student scores. However, teachers' median ratings of abstractness and overall demand were found to be significant predictors of the participating students' mean scores on the assessment items. Additionally, considering only the perceptions of the teachers who taught the five classes of students involved in this study, the teachers' ratings of complexity, strategy, and overall demand were significant predictors of their students' mean scores. The scores of the participating students also differed significantly between items with Level 1 and Level 4 ratings for the strategy dimension.

These results suggest that teachers' perceptions of overall demand provide a good indication of students' understanding and performance on an assessment item. This could be because when teachers gauged the cognitive demand of the items, they considered aspects such as students' prior exposure to the items and the difficulty of the mathematical concepts. In fact, the teachers correctly identified the areas in which they thought that students were likely to have difficulty, for example, associating $y = \mathrm{f}(x - a)$ with a translation of a units in the positive x-direction, and in understanding mathematical language and notations (e.g., quadratic polynomial, general term of a sequence). The teachers' perceptions of demand in terms of the three dimensions also produced modest success in predicting the difficulty experienced by students as evidenced by students' scores.

Conclusion

Teachers' perceptions of cognitive demand of mathematical assessment items are often not articulated. The findings from this study provide preliminary insights into what constitutes cognitive demand in A-level mathematics as perceived by teachers, and how they draw on their knowledge of both the subject and of their students to facilitate their judging of demand. The findings of this study also provide some indication that A-level mathematics teachers draw on their PCK in their process of judging cognitive demand. Although a limitation of this study is the small sample size, it adds to the body of evidence on the importance of CK and PCK in guiding mathematics teachers' instructional and assessment practices (Ball et al., 2008; Hoover, Mosvold, Ball, & Lai, 2016).

Further investigations could explore the interplay among the different dimensions of cognitive demand as perceived by teachers and the reasons for the differences in their perceptions. For example, how do teachers relate their SMK to the demand described in the cognitive demand dimensions? To what extent are teachers' perceptions of cognitive demand influenced by their knowledge of their own students? To date, there are few studies that have examined the impact of teachers' professional knowledge on teaching and learning as pointed out by Hoover et al. (2016). Thus, it could also be valuable to study the impact of teachers' professional knowledge on their perceptions of cognitive demand. It would also be meaningful to examine students' perceptions of cognitive demand to gain insights into the relationships between teachers' perceptions of cognitive demand as intended in the assessment items and the actual demand perceived by students.

References

Ball, D. L. (1990). The mathematical understandings that prospective teachers bring to teacher education. *The Elementary School Journal*, *90*(4), 449–466.

Ball, D. L., Lubienski, S. T., & Mewborn, D. S. (2001). Research on teaching mathematics: The unsolved problem of teachers' mathematical knowledge. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed., pp. 433–456). Washington, DC: American Educational Research Association.

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education*, *59*(5), 389–407.

Blömeke, S., & Delaney, S. (2012). Assessment of teacher knowledge across countries: A review of the state of research. *ZDM Mathematics Education*, *44*(3), 223–247.

Crisp, V., & Novaković, N. (2009). Is this year's exam as demanding as last year's? Using a pilot method to evaluate the consistency of examination demands over time. *Evaluation & Research in Education*, *22*(1), 3–15.

Gardner, H. (1985). *The mind's new science: A history of the cognitive revolution*. New York: Basic Books.

Goldhaber, D. D., & Brewer, D. J. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Educational Evaluation and Policy Analysis*, *22*(2), 129–145.

Gorin, J. S., & Embretson, S. E. (2013). Using cognitive psychology to generate items and predict item characteristics. In M. J. Gierl & T. M. Haladyna (Eds.), *Automatic item generation: Theory and practice* (pp. 136–156). New York: Routledge.

Grossman, P. L. (1990). *The making of a teacher: Teacher knowledge and teacher education*. New York: Teachers College Press.

Hoover, M., Mosvold, R., Ball, D. L., & Lai, Y. (2016). Making progress on mathematical knowledge for teaching. *The Mathematics Enthusiast*, *13*(1–2), 3–34.

Hughes, S., Pollitt, A., & Ahmed, A. (1998, August). *The development of a tool for gauging the demands of GCSE and A level exam questions*. Paper presented at the British Educational Research Association Annual Conference, Queen's University Belfast, UK.

National Governors Association Center for Best Practices, Council of Chief State School Officers (2010). *Common core state standards (Mathematics)*. Washington D.C.: Authors.

Pollitt, A., Ahmed, A., & Crisp, V. (2007). The demands of examination syllabuses and question papers. In P. Newton, J. A. Baird, H. Goldstein, H. Patrick, & P. Tymms (Eds.), *Techniques for monitoring the comparability of examination standards* (pp. 166–211). Retrieved from https://www.gov.uk/government/publications/techniques-for-monitoring-the-comparability-of-examination-standards

Porter, A., & Brophy, J. (1988). Synthesis of research on good teaching: Insights from the work of the institute for research on teaching. *Educational Leadership*, *45*(8), 74–85.

Rowan, B., Chiang, F.-S., & Miller, R. J. (1997). Using research on employees' performance to study the effects of teachers on students' achievement. *Sociology of Education*, *70*(4), 256–284.

Schoenfeld, A. H. (1985). *Mathematical problem solving*. Orlando, Fl: Academic Press.

Schoenfeld, A. H. (1988). When good teaching leads to bad results: The disasters of "well-taught" mathematics courses. *Educational Psychologist, 23*(2), 145–166.

Tan, H.C.J., Ng, W.L., & Shutler, P.M.E. (2017). Development and field-testing of an instrument for rating cognitive demands of mathematical assessment items. *The Mathematics Educator*, 17 (1&2), 57-78.

Van Driel, J. H., & Berry, A. (2009). The teacher education knowledge base: Pedagogical content knowledge. In P. Peterson, E. Baker, & B. McGraw, (Eds.), *International encyclopaedia of education* (3rd ed.; Vol. 7; pp. 656–661). London, GBR: Elsevier Science & Technology. Retrieved from http://www.ebrary.com

Verloop, N., Van Driel, J., & Meijer, P. (2001). Teacher knowledge and the knowledge base of teaching. *International Journal of Educational Research*, *35*(5), 441–461.

Authors:

Ng Wee Leng (**corresponding author:** weeleng.ng@nie.edu.sg),
Tan Hwee Chiat June (june.thc@gmail.com**)** &
Paul M. E. Shutler (paul.shutler@nie.edu.sg),
National Institute of Education, Nanyang Technological University,
1 Nanyang Walk, Singapore 637616.