International Journal of Assessment Tools in Education
2023, Vol. 10, No. 2, 175–196

https://doi.org/10.21449/ijate.1160120

Published at https://ijate.net/ https://dergipark.org.tr/en/pub/ijate **Research Article**

Validation of cognitive models for subtraction of time involving years and centuries

Huan Chin[1], Cheng Meng Chew[2,*]

[1]School of Educational Studies, Universiti Sains Malaysia, Penang, Malaysia
[2]School of Education, Humanities and Social Sciences, Wawasan Open University, Penang, Malaysia

ARTICLE HISTORY

Received: Aug. 10, 2022
Revised: May 01, 2023
Accepted: May 10, 2023

Keywords:
Cognitive models,
Cognitive diagnostic assessment,
Validation,
Attribute hierarchy method,
Learning of time.

Abstract: Years and Centuries are the measurement units used to quantify a longer time duration, while subtraction is the operation required to determine the duration based on two given time points. However, subtraction of time is a difficult skill to be mastered by many elementary students. To identify the root cause of the student's failure in performing subtraction involving the unit of time, we developed and validated the three cognitive models related to this skill by conducting a descriptive study which involved 119 Grade Five students from three Malaysian elementary schools. The cognitive diagnostic assessment developed based on the three cognitive models was used to elicit the participants' responses. Then, Attribute Hierarchy Method and Classical Test Theory were employed to analyse the data. The findings indicated that the hierarchical structures of all cognitive models are supported by the student's responses. The three student-based cognitive models were also highly consistent with the corresponding expert-based cognitive models. The cognitive models developed could guide diagnostic assessment development and diagnostic inference making.

1. INTRODUCTION

Time is one of the key concepts included in the domain of measurement in elementary school mathematics (National Council of Teachers of Mathematics [NCTM], 2000; Van de Walle et al., 2018). Time telling and determining the duration are the two key skills covered under the concept of time in elementary mathematics (Harris, 2008). Students learn about time telling in Grade One and Grade Two, followed by determining the duration of time in Grade Three onwards. Besides the commonly used time unit such as second, minute, hour, day, week, month, and year, students are also introduced to the year-based time unit such as year, decade, and century for quantifying a longer time duration. This numeracy capability will support the students in interpreting the historical timeline (Australian Curriculum, Assessment and Reporting Authority [ACARA], 2017).

In fact, subtraction of time is the mathematical operation which is needed to determine the duration between two given time points (Sia et al., 2019). Regardless of the time unit, the

*CONTACT: Cheng Meng Chew ✉ cmchew@usm.my ; cmchew@wou.edu.my School of Educational Studies, Universiti Sains Malaysia, Penang, Malaysia & School of Education, Humanities and Social Sciences, Wawasan Open University, Penang, Malaysia

e-ISSN: 2148-7456

students would have to subtract the began time (subtrahend) from the ended time (minuend) to find the duration. Thus, the students were exposed to the procedural skills exercise related to subtraction of time as shown in Figure 1(a) prior proceed with the mathematical task involving duration as shown in Figure 1(b) (Chan et al., 2017). Specifically, the students would perform subtraction involving time by using the column method.

Figure 1. *Mathematical task related to 'Subtraction of Time' which involves Centuries and Years. Adopted from Chan (2017).*

2 centuries 48 years − 84 years =

Column method:

	centuries	years
	2	48
−		84

(a) Procedural Skills Exercise

The table below shows the time taken by two planets to make one complete revolution around the Sun.

Planet	Duration
Pluto	2 centuries 48 years
Uranus	8 decades 4 years

Calculate the difference in the time taken, in years, by Pluto and Uranus to make one complete revolution around the Sun.

(b) Mathematical Task involving Duration

Even though procedural fluency has been emphasized in the mathematics classroom, students' failure in subtraction of time is frequently highlighted by researchers (i.e., Earnest, 2015; Kamii & Russell, 2012; Sia et al., 2019). Despite subtraction and conversion of time have been introduced to the students in the early grade, the students seem to fail to integrate their knowledge about subtraction and the relationship between the time units for performing subtraction involving the unit time. In this regard, cognitive modelling could be an appealing approach to deepen the educators' understanding of the student's failure in performing subtraction involving the unit of time from the perspective of their cognition (Leighton & Gierl, 2007).

The use of cognitive models in supporting the process of making diagnostic inferences has been illustrated in the studies conducted by Leighton and Gierl (2007), and Sia et al. (2019). While cognitive models were used to guide the test items' construction, students' performance would be tapped into mastery of the subskill which was arranged hierarchically as illustrated in the cognitive models. Thus, the use of cognitive models in test development and result interpretation would eventually highlight the students' cognitive weaknesses which hinder their mastery of the concept and hence explain their performance (Kane & Bejar, 2014). With the rich diagnostic information, the teachers could differentiate the teaching and plan for remedial instruction to support students in learning the subtraction of 'Time'. Thus, this study sought to develop and validate the cognitive models for 'Subtraction of Time' involving *years* and *centuries*.

1.1. Literature Review

1.1.1. *Cognitive models*

In the field of cognitive psychology, the cognitive model refers to the theoretical foundation of the procedures which are assumed to be carried out during complex cognitive activities such as problem-solving and decision-making (Keehner et al., 2017). The integration of cognitive psychology and educational measurement brings about the emergence of two types of cognitive models, namely *cognitive models of learning* and *cognitive models of task*. According to Keehner et al. (2017), the *cognitive models of learning* such as learning progressions describe the stages of knowledge and skill acquisition as well as competence development. Meanwhile, the *cognitive models of task* refer to the descriptions of the attribute used by the students in solving some tasks, for supporting the inferences made on students' performance (Leighton & Gierl, 2007).

Assessment practice can be implemented with the integration of both *cognitive models of task* as well as *cognitive models of learning*. By applying the specific cognitive-psychometric model, students' responses elicited using the test items guided by the *cognitive models of learning* will be mapped onto the corresponding developmental stage as stated in the *cognitive models of learning*. Meanwhile, fitting students' responses prompted using test items guided by *cognitive models of task* onto the cognitive-psychometric model would generate students' attribute mastery. In order to ensure the diagnostic information generated based on the cognitive models is highly precise and specific, the cognitive model of task must possess four significant properties: (i) each fine-grained attribute must be specified in a detailed manner consistently; (ii) each attribute must be able to measure using cognitive task; (iii) each attribute specified must be aligned with the curriculum; and (iv) each attribute must be structured hierarchically in the cognitive model (Gierl et al., 2009b).

The previous relevant studies mainly focused on the development and validation of *cognitive models of learning* for various topics of science, such as ecosystem (Jin et al., 2019), matter (Hadenfeldt et al., 2016), and phase transformation (Schultz et al., 2017). For the subject of mathematics, the cognitive models of learning have been developed to describe the developmental stages of a concept such as subtraction of fractions (Akbay et al., 2018) as well as number sense (Chen et al., 2017). Rather than focusing on the validation of *cognitive models of learning*, Sia et al. (2019) evaluated the consistency of the *cognitive model of task* for the '*duration'* concept which was hypothesized by the expert, and the *cognitive model of task* exhibited on students' actual cognitive processes when solving the tasks.

In this study, we focused on the development and validation of the cognitive models of task which can be used to support the diagnostic inference made regarding students' cognitive strengths and weaknesses which contribute to the mastery or non-mastery of the skills in 'Subtraction of Time'. We validated the *cognitive models of task* by using various sources of evidence to validate the data, rather than just evaluating the consistency as demonstrated by Sia et al. (2019). The term '*cognitive model*' will be used to indicate the '*cognitive model of task*' in this study.

1.1.2. *Development of cognitive models*

Ideally, the cognitive models should be constructed based on a substantive theory of cognition and learning (Nichols, 1994). Since a suitable substantive theory can rarely be found in the literature, Gierl et al. (2009a) introduced two approaches for developing the cognitive model, namely the top-down approach and the bottom-up approach. The top-down approach involves conducting a task analysis within the domain of interest for developing the cognitive models, whereas the bottom-up approach involves analysing the protocol data collected using the think-aloud method for developing the cognitive models.

The use of the top-down approach has been demonstrated in the study conducted by Gierl et al. (2008), as well as Sia and Lim (2018). They began with specifying the attributes which include the mathematical concepts, skills, and processes used to solve each task, followed by arranging those attributes hierarchically based on their complexity to form the cognitive models. Instead of conducting the task analysis, Chen et al. (2017) identified the attributes by reviewing the skills included in the textbooks.

Meanwhile, the use of the bottom-up approach has been demonstrated in the study conducted by Gierl et al. (2009a). They started with transcribing the recordings, coding the attribute, and presenting the process to solve the task by using the flowchart which represents the cognitive models from the student's perspective. Notably, the top-down approach has been used in majority of the previous studies due to the advantages of this approach in ensuring the cognitive models' credibility. With sufficient teaching experience, the experts would have a strong understanding of students' thinking, learning and instruction. This relevant expertise would eventually support them in identifying and arranging the instructional relevant attributes in hierarchical order to form the cognitive models (Gierl et al., 2009b).

1.1.3. *Validity of cognitive models*

As asserted by Leighton and Gierl (2007), the substantive theory of learning and cognition integrated into the assessment is in high demand. Thus, the cognitive models constructed are regarded as a new theory of learning and cognition (Nichols et al., 2017) which has never been demonstrated to be valid (Nichols, 1994). In other words, the attributes associated with students' cognitive processes in solving the related tasks are only hypothesized by the experts (Graf et al., 2019). Thus, empirical evidence needs to be accumulated to verify the validity of the cognitive models constructed (Graf et al., 2019).

The empirical evidence can be collected from various sources based on the nature of constructs and tasks. Most of the studies (i.e., Akbay et al., 2018; Chen et al., 2017; Graf et al., 2019; Langenfeld et al., 2020; Sia et al., 2019) employed cognitive-psychometric modelling to validate the cognitive models because the fit of cognitive models to the data can be computed by using the mathematics formula which take into consideration of the constraints imposed in the cognitive models (Keehner et al., 2017). In order to triangulate the data, Graf et al. (2019), Langenfeld et al. (2020) and Sia et al. (2019) further analysed the protocols collected by using the think-aloud method and conducting cognitive labs, respectively.

Unlike the study conducted by Langenfeld et al. (2019) and Sia et al. (2019), we began the empirical evaluation of cognitive models by conducting Item Difficulty Modelling (IDM) as suggested by Keehner et al. (2017). Following this, we could explore the cognitive models' hierarchical structure by taking into account the item difficulty (Gorin, 2006) and attribute complexity (Keehner et al., 2017). After conducting IDM, we evaluated the consistency between the student-based cognitive models (S-CM) and expert-based cognitive models (E-CM) by applying the cognitive-psychometric modelling.

1.1.4. *The learning of subtraction concept*

The concept of subtraction is commonly introduced to children informally (Clement et al., 2020) using the 'taking away', 'part-part-whole' and 'comparing two sets of items' problem situations (Carpenter et al., 1999). After conceptualising subtraction operation, the subtraction learning will begin with single-digit numbers with a unitary conceptual structure (Fuson, 1990). The students will engage in solving simple subtraction problems involving single-digit numbers (Clement et al., 2020) using Count-All-and Taking-Away (Murata & Kattubadi, 2012). After that, the students will be guided to make use of the numerical information at the subtrahend to find the difference using the Counting-Up or Counting Down strategy (Murata & Kattubadi, 2012). Once the students understand the relationship between minuend and subtrahend, they will be introduced to find the difference using the subtraction algorithm (Murata & Kattubadi, 2012).

The learning of subtraction is then extended to multi-digit numbers. The learning of multi-digit subtraction is more complex because the multi-digit number is conceptualized as 'multiunit quantities associated with multiunit names and position (Fusion, 1990, p. 350)'. For example, the two-digit number '23' is conceptualized as a combination of two bundles of 10 sticks and three single sticks. Thus, the understanding of the base-ten place value system preceded multi-digit subtraction (Fuson, 1990; Nuerk et al., 2015). Following this, Nuerk et al. (2015) suggested that multi-digit subtraction involves three processes: (i) place identification, (ii) place-value activation, and (iii) place-value computation. It begins with assigning each digit to the correct base-10 place value stack position (*place identification*), followed by conceptualising the digit located at the respective base-10 place value stack position as corresponding numerical magnitude (*place-value activation*), and regrouping the numbers and performing the subtraction across place-value stacks (*place-value computation*) (Nuerk et al., 2015).

The measurement of time involving a composite unit can be expressed as a multi-unit conceptual structure because it involves the pairing of a number associated with a higher measurement unit and a number associated with a lower measurement unit (Fusion, 1990b). Even though the higher measurement unit quantifies a collection of the lower measurement unit, the value of the number associated with the higher measurement units might not always equal 10 times the number associated with the lower measurement units. Thus, subtraction involving measurement might be slightly different from multi-digit subtraction which follows the base-10 place value system.

In fact, the learning of subtraction involving the measurement of time can be extended from multi-digit subtraction using different number bases considering the relationship between the units of time. In this study, learning of subtraction involving century and year could be extended from multi-digit subtraction with a number base of 100 because a century is quantified as 100 years.

1.1.5. *Past related studies on time concept*

While time telling and duration of time are two fundamental concepts in the topic of 'Time', several studies have been conducted on assessing students' performance on time telling (Brace et al., 2019; Lambert et al., 2020), as well as identifying students' common errors (Tan et al., 2019) and knowledge state (Tan et al., 2017) in finding duration of calendar time. Besides that, past studies (e.g., Chin et al., 2021b, 2022) also focused on developing assessments to measure students' attribute mastery level for addition and multiplication of time involving *hours, days, weeks, months, and years.* Meanwhile, several interventions have been introduced by the researchers to support students' learning of time-telling (Earnest, 2021; Pelton et al., 2018; Wang et al., 2016) as well as the learning of time concepts involving *years, decades, and centuries* (Chin et al., 2021a).

For the aspect of subtraction of time, the past studies mainly focused on error analysis (i.e., Earnest, 2015; Kamii & Russell, 2012; Ojose, 2015). The findings constantly reported that most of the students tend to make regrouping errors when performing subtraction of time. For instance, they tend to subtract the measurement of time without performing necessary regrouping (Kamii & Russell, 2012). Even though some of the students understand the concept of regrouping, Earnest (2015) and Ojose (2015) found that the students tend to make mistakes in regrouping the time notation. Besides confusing the time notation with the base ten number system (Earnest, 2015), the students also regrouped the time notation using the wrong time relationship (Ojose, 2015). For example, the students might regroup 1 day into 12 hours rather than 24 hours.

1.1.6. *The present study*

'Subtraction of Time' is regarded as a difficult skill to be mastered by students. Despite the importance of subtraction of time involving years and centuries in determining a longer duration between two given time points, the previous relevant studies mainly concentrated on the frequently used time unit, such as *hours* and *minutes* (e.g., Earnest, 2015; Kamii & Russell, 2012; Sia et al., 2019). The study focused on the subtraction of time involving *years*, and *centuries*, which are rarely found in the literature (Chin et al., 2021a).

To determine the persistence made by the students in subtraction involving time, several studies (i.e., Earnest, 2015; Kamii & Russell, 2012; Ojose, 2015) have been conducted by performing error analysis. However, this approach fails to pinpoint the underlying cognitive attribute deficit which leads to the errors made (Ketterlin-Geller & Yovanoff, 2009). Consequently, the students' procedural errors were usually corrected without considering the conceptual understanding (Russell & Masters, 2009) which provides strong support for the development of procedural knowledge and permits the extension of the mathematical idea (Rittle-Johnson & Schneider, 2015).

In this regard, the use of cognitive models in test development and result interpretation could provide informative diagnostic data for supporting the teachers in planning the remedies to support students' acquisition of 'subtraction of time'. Yet, the available cognitive models which can be used to support the highly specific diagnostic inference made are rarely found in the literature (Gierlet al., 2009a; Sia et al., 2019). Besides that, the cognitive process of performing subtraction of time involving years and centuries is left unexplored in the past.

To fill the research gap, this study sought to develop and validate the cognitive models for 'Subtraction of Time' involving years and centuries. In this paper, we present the process of developing the cognitive models. To ensure the validity of the diagnostic claims made, we validate the cognitive models developed by addressing the following research questions:

(1) To what extent are hierarchical arrangement of attribute in expert-based cognitive models supported by students' responses?
(2) To what extent are the attribute dependency in the expert-based cognitive models supported by students' responses?
(3) To what extent are the expert-based cognitive models consistent with the student-based cognitive models for 'Subtraction of Time'?

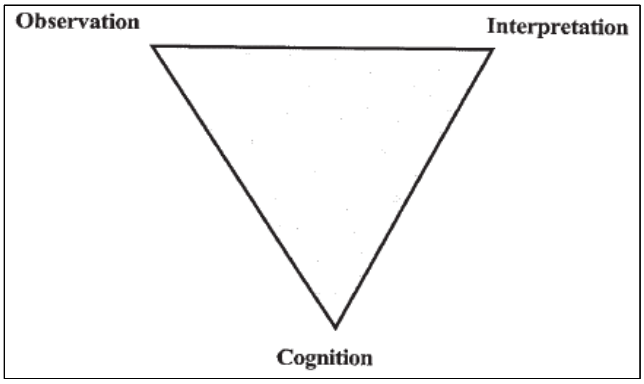
1.2. Theoretical Framework

1.2.1. *Assessment triangle*

The development and validation of cognitive models were grounded in the framework called Assessment Triangle (Pellegrino et al., 2001). This framework explains the mechanism of linking educational measurement with human cognition by using a triangle as illustrated in Figure 2. The three basic elements of assessment, namely (i) a cognitive model which illustrates the students' skills or knowledge acquisition in the tested domain, (ii) the task which triggers students' response to manifest their skills or knowledge and (iii) the interpretation method used to make diagnostic inferences (Pellegrino & Chudowsky, 2003) are pivoted on each vertex of the Assessment Triangle namely, cognition, observation, and interpretation, respectively. Following this, each element is linked to the other two elements and works in synchrony. Hence, students' cognition can be used to explain their strengths and weaknesses (Pellegrino et al. 2001).

The cognitive models which are embedded in the cognition vertex of the Assessment Triangle were developed by using the top-down approach in this study. Since the attributes which form the cognitive models are considered latent traits that are non-observable (Keehner et al., 2017), the assessment tasks were developed to elicit students' responses which could demonstrate their mastery of attributes. These assessment tasks are pivoted to the observation vertex of the Assessment Triangle. While the cognitive models represent the hierarchically ordered attributes, fitting the cognitive-psychometric model in the interpretation vertex, such as Attribute Hierarchy Method (AHM) onto students' responses collected in the observation vertex can be used to validate the cognitive models. (Leighton et al., 2004).

Figure 2. *Assessment triangle (Pellegrino & Chudowsky, 2013, p. 112).*



2. METHOD

A descriptive research design was adopted for the empirical validation of the cognitive models developed, which is predominantly descriptive. In this section, we discuss the process of developing the cognitive models, followed by describing the participants, research instruments, and the research procedure of the empirical validation of the cognitive models constructed.

2.1. Participants

Since the development of the cognitive model is an iterative process, the participants of the study were selected by employing convenience sampling, which is commonly used for piloting an under-developed instrument (Salkind, 2010) because of its main advantage in terms of cost efficiency. A total of 119 Grade Five students from one National School (NS), one National-Type Chinese School (NTCS) and one National-Type Tamil School (NTTS) in Penang, Malaysia with the Malay, Mandarin, and Tamil language as the medium of mathematics instruction, respectively were chosen to participate in the study. Since class streaming is no longer practised in the schools, each class consisted of students with mixed abilities. In order to ensure the representativeness of the data, the intact class of the students were chosen. With the sample size of 119, which surpassed the minimum sample size required (n=100) for employing the psychometric model named AHM, which is rooted in latent class analysis (Wrupts & Geiser, 2014), the findings of the study could be reliable.

2.2. Development of Expert-Based Cognitive Models

The development of expert-based cognitive models started with the identification of attributes through task analysis and expert reviews following the suggestions were given by Gierl et al. (2010). A workshop which involved two invited mathematics education experts was conducted to specify the fine-grained attributes through task analysis and expert reviews in a workshop. The background of the two experts is presented in Table 1.

Table 1. *Background of mathematics education experts.*

Expert	Academic Qualification	Specialization	Position	Affiliation
Expert 1	Doctor of Philosophy	Mathematics Education	Associate Professor	Public University in Malaysia
Expert 2	Doctor of Philosophy	Mathematics Education	Associate Professor	Public University in Malaysia

During the workshop, the two experts reviewed Year Four Mathematics Textbook as well as the Curriculum and Assessment Standard Document to deepen their understanding of the learning standards related to the intended construct, that is 'Subtraction of Time'. Then, they listed the main skills about 'Subtraction of Time' as tabulated in Table 2 based on the learning standards.

Table 2. *Main skills related to 'Subtraction of Time'.*

Main Skill	Description
Main Skill 1	Subtraction of time involving century and year without regrouping
Main Skill 2	Subtraction of time involving century and year with single regrouping
Main Skill 3	Subtraction of time involving century and year with double regrouping

After that, task analysis was performed by the two experts on the task selected from the textbook based on each main skill as demonstrated in Figure 3. During the task analysis, each step involved in solving the tasks is depicted as a detailed description. Based on this description, the experts outlined the attributes which can be measured using the test items (Alves, 2012). For example, the attribute *'Convert 1 century to 100 years, add the 100 years into the number of years in the first minuend and subtract the number of centuries and years in the subtrahend from the first minuend'* is used to summarize the description of steps: (i) '*Borrow 1 century from the century column in the first minuend*'; (ii) '*Convert 1 century to 100 years and add the 100 years into the number of years in the first minuend';* and (iii) '*Subtract the number of centuries and years in the first subtrahend from the first minuend*'. Notably, this attribute could barely be measured using an item because it is less precise. Thus, it was rephrased into a clearer version such as '*Subtract one unit of time from one unit of time involving century and year with regrouping*'. The modified attributes are shown in the square brackets in Figure 3.

To ensure the instructional relevance of the attributes, a panel of subject matter experts (SMEs) were invited to validate the attributes. The background of the SMEs is shown in Table 3. All attributes were rated as '5' by each pair of experts on the 5-point Likert-scale validation form. With the simple agreement of 100 percent, the relevancy of the attributes with respect to the content standards and learning standards was very high.

Table 3. *Background of the subject matter experts.*

Expert	Academic Qualification	Specialization	Position	Affiliation
Expert 1	Master in Education	Mathematics Education	Experienced Teacher	NS in Malaysia
Expert 2	Master in Education	Mathematics Education	Experienced Teacher	NS in Malaysia
Expert 3	Master in Education	Mathematics Education	Experienced Teacher	NTCS in Malaysia
Expert 4	Master in Education	Mathematics Education	Experienced Teacher	NTCS in Malaysia
Expert 5	Master in Education	Mathematics Education	Experienced Teacher	NTTS in Malaysia
Expert 6	Doctor of Philosophy	Mathematics Education	Experienced Teacher	NTTS in Malaysia

After the validation process, the attributes were ordered hierarchically by the two mathematics experts, based on the complexity of each attribute to derive the attribute hierarchy. For example, the attribute *‘Subtract one unit of time from one unit of time involving century and year with regrouping’* was positioned at the lowest level of the hierarchy since it is less complex compared to the attribute *‘Subtract two units of time from one unit of time involving century and year with single regrouping’*. A total of six attribute hierarchies related to ‘Subtraction of Time’ as shown in Table 4 were specified by the experts in the field of mathematics education. These attribute hierarchies are considered as Expert-based Cognitive Models (E-CM).

Table 4. *Three attribute hierarchies related to ‘Subtraction of Time’.*

Cognitive Model	Attribute Hierarchy	Attributes
Cognitive Model 1	CM1A1 → CM1A2	CM1A2: Subtract two units of time from one unit of time involving century and year without regrouping. CM1A1: Subtract one unit of time from one unit of time involving century and year without regrouping.
Cognitive Model 2	CM2A1 → CM2A2	CM2A2: Subtract two units of time from one unit of time involving century and year with single regrouping. CM2A1: Subtract one unit of time from one unit of time involving century and year with regrouping.
Cognitive Model 3	CM3A1 → CM3A2	CM3A2: Subtract two units of time from one unit of time involving decade and year with double regrouping. CM3A1: Subtract one unit of time from one unit of time involving century and year with regrouping.

Figure 3. *Task analysis for Main Skill 2.*

Working	Description of the steps	Attributes	Attribute Hierarchy
Century / Year ① 7 ~~8~~ / 100+6 = 106 ② ~~6~~ − 2 / 7 ③ 5 / 99 − 9 ④ 5 / 90	① Borrow 1 century from the decade column in the first minuend. ② Convert 1 century to 100 years and add the 100 years into the number of years in the first minuend. ③ Subtract the number of cenrturies and years in the first subtrahend from the first minuend.	**CM2A1**: Convert 1 century to 100 years, add the 100 years into the number of years in the first minuend and subtract the number of centuries and years in the subtrahend from the first minuend. [Subtract one unit of time from one unit of time involving century and year with regrouping]	CM2A1 → CM2A2
	④ Subtract the number of centuries and years in the second subtrahend from the first minuend.	**CM2A2**: Subtract the number of centuries and years in second subtrahend from the second minuend. [Subtract two units of time from one unit of time involving century and year with single regrouping]	

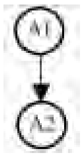
To ensure the appropriateness of the sequence of the attributes being ordered, these attribute hierarchies underwent validation that involved six subject matter experts. All attribute hierarchies were rated as '5' by each pair of experts on the 5-point Likert-scale validation form. With the simple agreement of 100 percent, the arrangement of the attribute hierarchies was considered very appropriate. These validated attribute hierarchies were regarded as expert-based cognitive models (Sia et al., 2019).

In order to validate the cognitive models, a matrix-formed test specification, named reduced Q matrix (**Q_r** matrix), was derived from the expert-based cognitive models. The hypothesized attributes that need to be mastered in order to answer each test item correctly are depicted in the reduced Q-matrix (Li & Suen, 2013). The **Q_r** matrix derivation process is illustrated in Figure 4. The derivation of **Q_r** matrix began with using the second order binary square matrix named *adjacent matrix* (**A** matrix) to specify the direct attributes' relationship in the hierarchy. In the A matrix, the presence or absence of the direct attributes' relationships was represented using '1' or '0' respectively. Then, the direct and indirect attributes relationships in the hierarchy were represented using the second order binary square matrix, named *reachability matrix* (**R** matrix) derived by applying Boolean arithmetic following the formula $\mathbf{R} = (\mathbf{A} + \mathbf{I})^n$, where **I** refers to the *Identity matrix* and *n* refers to the smallest integer needed to obtain a constant **R** matrix. After that, the number of potential items (*i*) was determined by employing the formula $i = 2^k - 1$, where *k* indicates the number of attributes.

Then, the incidence matrix (**Q** matrix) was derived to portray the attribute combinations which might be involved in solving each potential item correctly. Then a further derivation was conducted to reduce the **Q** matrix of order 2 × 3 into a binary second order squared matrix named reduced incidence matrix (**Q_r** matrix) by establishing the direct and indirect attribute relationships following its specification shown in the **R** matrix. For instance, the removal of the second column of **Q** matrix was made due to the fact that the items that involve attribute CM2A2, would also involve attribute CM2A1 indirectly. In other words, none of the items could be used to measure solely the attribute CM2A2, without measuring attribute CM2A1 indirectly. After deriving the **Q_r** matrix, the cognitive diagnostic assessment (CDA) can be constructed to collect the empirical data for validating the expert-based cognitive models developed.

Figure 4. *The derivation of the Qr matrix (Chin et al., 2021b, p. 300).*

Attribute Hierarchy



- The attributes A1, and A2 are arranged in a linear hierarchical order based on their complexity to form the attribute hierarchy.

A matrix

$$\begin{array}{c} \\ A1 \\ A2 \end{array} \begin{array}{c} \begin{array}{cc} A1 & A2 \end{array} \\ \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \end{array}$$

- ***A*** matrix is used to represent the direct relationship of the attributes in cognitive diagnostic assessment (Tatsuoka 1986).
- The presence and absence of the direct relationship between attributes are represented as ‘1’ and ‘0’ respectively in the ***A*** matrix.

R matrix

$$\begin{array}{c} \\ A1 \\ A2 \end{array} \begin{array}{c} \begin{array}{cc} A1 & A2 \end{array} \\ \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \end{array}$$

- ***R*** matrix is used to represent the direct and indirect prerequisite relationship among the attributes in cognitive diagnostic assessment (Tatsuoka 1986).
- The presence or absence of the direct and indirect relationship between attributes is represented as ‘1’ and ‘0’ respectively in the ***R*** matrix.
- ***R*** matrix can be derived from ***A*** matrix by performing Boolean addition and multiplication using formula $R = (\boldsymbol{A}+\boldsymbol{I})^{n}$, where ***I*** refers to the identity matrix and n refers to the integer required to reach invariance.

Q matrix

$$\begin{array}{c} \\ A1 \\ A2 \end{array} \begin{array}{c} \begin{array}{ccc} Q1 & Q2 & Q3 \end{array} \\ \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \end{array}$$

- The incidence matrix (**Q** matrix) is the potential pool of items which represent all combination of attributes when the attributes are assumed to be not related to each other.
- The number of potential items can be determined using the formula
  $i = 2^{k}-1$, where k is the number of attributes

$\boldsymbol{Q_r}$ matrix

$$\begin{array}{c} \\ A1 \\ A2 \end{array} \begin{array}{c} \begin{array}{cc} Q1 & Q3 \end{array} \\ \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \end{array}$$

- According to Gierl, Leighton, and Hunka.(2000), the reduced incidence matrix ($\boldsymbol{Q_r}$) represents the items from the potential pool that satisfy the attribute hierarchy as shown in the cognitive model.
- ***Q*** matrix can be reduced to form the $\boldsymbol{Q_r}$ matrix by imposing the prerequisite relation of the attributes as specified in the ***R*** matrix (Gierl et al. 2000; Tatsuoka 1991, 2009).

2.3. Research Instrument

The validity evidence of the cognitive models was collected by using the CDA for 'Subtraction of Time' which was developed based on the $\mathbf{Q_r}$ matrix. Each combination of attributes depicted at each column of the $\mathbf{Q_r}$ matrix was probed using three parallel open-ended items as recommended by Sia and Lim (2018) to enhance the reliability of the CDA (Gierl et al., 2009). Following this, the CDA for 'Subtraction of Time' consisted of 18 items to elicit student responses for the two attribute combinations as shown in each $\mathbf{Q_r}$ matrix of the six cognitive models (2 cognitive models × 2 attribute combinations × 3 parallel items = 18 items). Corresponding to the six cognitive models, the CDA consisted of 3 sections as listed in Table 5. These English-written items were then translated into Malay, Mandarin and Tamil languages to comply with the instruction medium of mathematics lessons in NS, NTCS and NTTS, respectively.

Table 5. *Content of CDA.*

Section	Cognitive Model	Skill	Number of Items
Section A	Cognitive Model 1	Subtraction of time involving century and year with no regrouping	6
Section B	Cognitive Model 2	Subtraction of time involving century and year with single regrouping	6
Section C	Cognitive Model 3	Subtraction of time involving century and year with double regrouping	6
		Total	18

To ensure the content validity of the CDA, the two subject matter experts with at least 10 years of teaching experience each from NS, NTCS and NTTS were invited to validate the instrument after the translation process. All items were rated as '5' by the six experts on the 5-point Likert-scale validation form. With the content validity index of 1.00 at the scale level, all items in the CDA were highly relevant with respect to the corresponding attribute combination measured (Polit & Beck, 2006). After the validation process, the CDA was piloted using 32 Year Five NS pupils, 35 Year Five NTCS pupils and 15 Year Five NTTS pupils selected through convenience sampling, for determining the reliability of the instrument. Although the CDA comprised a set of open-ended items, it was scored dichotomously in accordance with the use of AHM as the psychometric model (Wang & Gierl, 2011). With the reliability coefficient of .90 which was calculated using Kuder Richardson 20, the dichotomously scored open-ended CDA was reliable (Multon & Coleman, 2010).

2.4. Research Procedures

The CDA was administered to the 119 Grade Five students in one NS, one NTCS, and one NTTS located in Penang, Malaysia. No time limit was imposed upon the test because the CDA was not served for students' performance comparison. After the test administration, the answer scripts were scored dichotomously. For each item, one mark was awarded to the correct response, while no mark was awarded to the incorrect response.

The pupils' responses were then further analysed by applying AHM. Specifically, Artificial Neural Network (ANN) pattern recognition analysis (PRA) was performed using Statistical Package for Social Sciences (SPSS) version 24 to estimate the pupil's attribute probability that corresponded with their response patterns in CDA. The ANN PRA is a two-stage data analysis process. During the first stage, the training of ANN was conducted so that the expected response pattern (ERP) could be associated with the corresponding expected attribute pattern (EAP) as shown in Table 6. To prevent the arisen of the issue regarding model-underfit, the ANN training data set consisted of the data made up of 100 replications of each ERP Vector and the corresponding EAP Vector pairs (Briggs & Kizil, 2017) as shown in Table 6. Following this, the ANN was trained with 300 samples (3 ERP-ERP pairs × 100 times of replication = 300 samples) by using the gradient transient backpropagation algorithm (Cui et al., 2016). With the architecture of 6 input nodes, 2 hidden nodes, and 2 output nodes, the error of the ANN converged at nearly zero (Root Mean Squared Error = .0088), which is acceptable (Cui et al., 2016). This indicates the relationship between the ERP Vectors and the corresponding EAP Vectors has been established. Following this, the pupils' attribute probabilities were estimated using the trained ANN at the second stage of ANN PRA.

Table 6. *Expected response pattern and expected attribute pattern.*

Expected Response Pattern	Expected Attribute Pattern
[0 0 0 0 0 0]	[0 0]
[1 1 1 0 0 0]	[1 0]
[1 1 1 1 1 1]	[1 1]

To determine the hierarchical arrangement of the attributes in the student-based cognitive models, item difficulty modelling was performed for each section of the assessment. Since the sample size of the study failed to meet the minimum requirement for performing Rasch Analysis (n=250) suggested by Linacre (1994), the item difficulty of each item was computed following the Classical Test Theory. Then, the position of each attribute in the hierarchy was determined based on the mean difficulty of the item measuring each attribute. Meanwhile, the mean attribute probabilities were used to confirm the hierarchical arrangement of the attributes in the S-CM. Then, the dependency of the attributes in the cognitive models was determined by referring to the correlation between the attribute probabilities. After portraying the S-CM, the Hierarchical Consistency Index (HCI) for each cognitive model based on the formulae as shown below, by using Microsoft Excel 2016 to determine the extent to which the S-CM are consistent with the E-CM.

$$HCI_i = 1 - \frac{2\sum_{j\in S_{correct_i}} \sum_{g\epsilon S_j} X_{i_j}(1 - X_{i_g})}{N_{c_i}}$$

Where,

$S_{correct_i}$ is a set which consists of the items that are correctly answered by the student *i*

X_{i_j} is the score (1 or 0) of student *i* for the item *j*, where item *j* is an element in the set $S_{correct_i}$

S_j is a set which consists of the items which required subset of attributes measured by item *j*, where item $j \notin S_j$

X_{i_g} is the score (1 or 0) of student *i* for the item *g*, where item *g* is an element in the set S_j

N_{c_i} is the total number of comparisons for all the items that are correctly answered by the student *i*

(Cui & Leighton, 2009, p. 436)

3. FINDINGS

3.1. Hierarchical Arrangement of Cognitive Models

Item difficulty modelling was used to verify the position of attributes in the cognitive models based on student responses. The item difficulty index, attribute-level mean item difficulty, and attribute-level z-score was computed for dispersion of the data. As shown in Table 7, the attribute level z-scores ranged from -1.11 to 1.15. This indicates that the difficulty of each item only differs by nearly one standard deviation of the respective attribute-level mean item difficulty. In other words, the three parallel items measuring the same attribute were calibrated at almost the same level of difficulty. As shown in Table 7, the p-value range of the first three items did not overlap with the p-value range of the last three items in each section. In other words, the items which measured the same attribute were grouped into a cluster calibrated with different complexity. This suggests the existence of a clear distinction in terms of the complexity between each attribute pair (i.e., CM1A1-CM1A2, CM2A1-CM2A2, and CM3A1-CM3A2) in the cognitive models. Thus, the two attributes in each cognitive model were not at the same level of hierarchy.

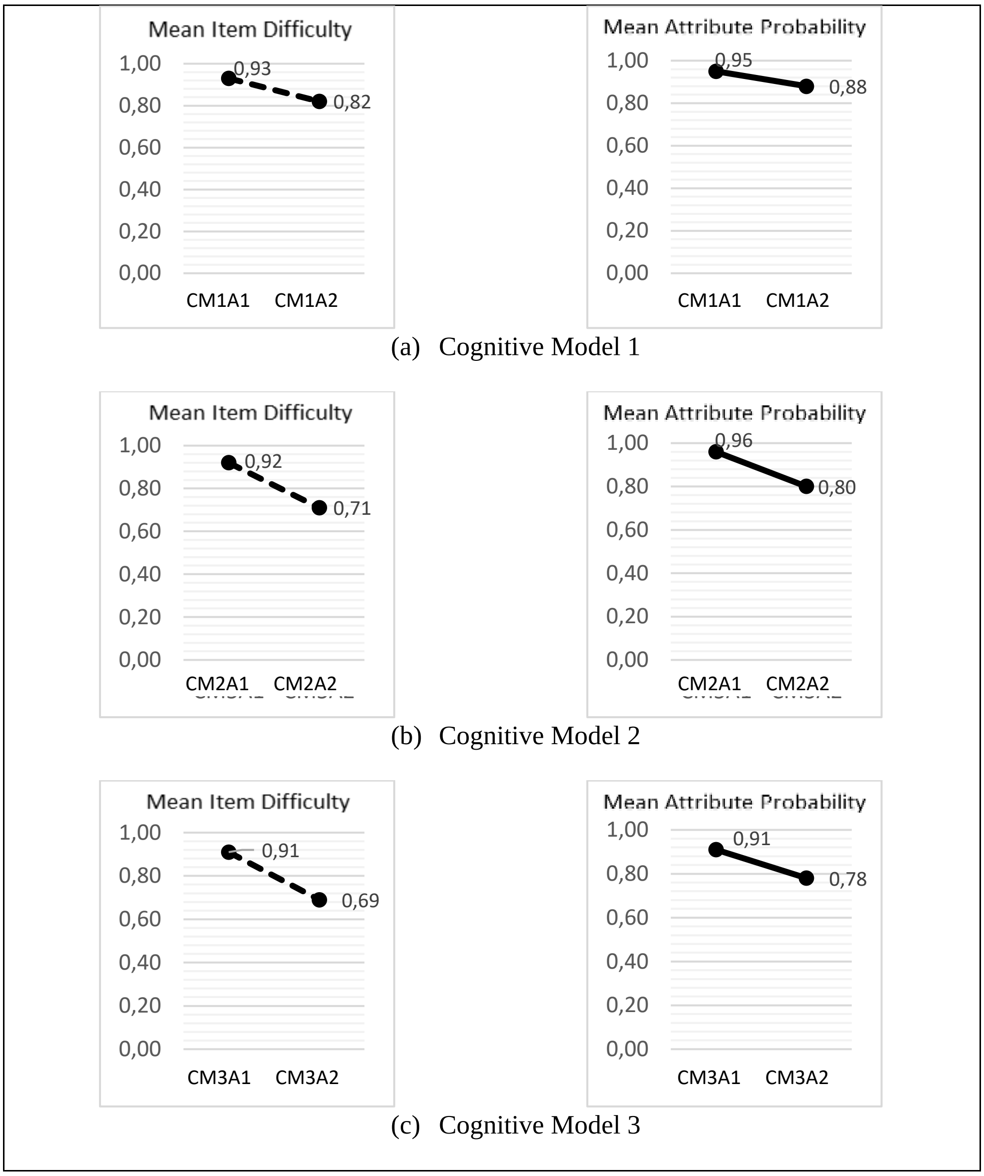
Table 7. *Item difficulty index, attribute-level mean item difficulty, and attribute-level z-score.*

Section [Cognitive Model]	Item	Item Difficulty Index (p-value)	Attribute Level Mean Item Difficulty Index	Attribute Level Z-score
Section A [Cognitive Model 1]	CM1A1		0.93	
	Item 1	0.94		1.15
	Item 2	0.92		-0.58
	Item 3	0.92		-0.58
	CM1A2		0.82	
	Item 4	0.84		1.06
	Item 5	0.84		1.06
	Item 6	0.79		1.00
Section B [Cognitive Model 2]	CM2A1		0.92	
	Item 1	0.89		-1.11
	Item 2	0.93		0.28
	Item 3	0.95		0.83
	CM2A2		0.71	
	Item 4	0.77		1.32
	Item 5	0.77		1.32
	Item 6	0.58		1.00
Section C [Cognitive Model 3]	CM3A1		0.91	
	Item 1	0.92		1.00
	Item 2	0.91		0.00
	Item 3	0.89		-1.00
	CM3A2		0.70	
	Item 4	0.74		1.06
	Item 5	0.66		0.94
	Item 6	0.70		1.00

Then, the attribute level mean item difficulty was compared for each section in order to determine the hierarchical position of the attributes in each cognitive model. The attribute level mean item difficulty was tabulated in Table 7. For each cognitive mode, the first attribute (i.e., CM1A1, CM2A1, and CM3A1) has a higher mean item difficulty compared to the second attribute (i.e., CM1A2, CM2A2, and CM3A2). This implies more students answered the items that probe the first attribute in each cognitive model correctly on average. In other words, the first attribute of each cognitive model is more basic than the second attribute of each cognitive model. Thus, the first attribute of each cognitive model was placed at a lower position in the hierarchy structure, compared to the second attribute in each cognitive model.

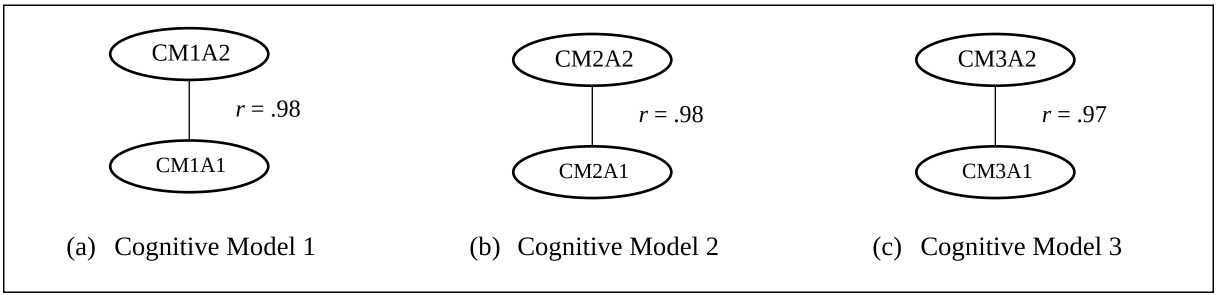
Then, the comparison of the mean item difficulty and mean attribute probability trend was made to further confirm the cognitive models' hierarchical structure. The results of the comparison were illustrated in Figure 5.

Figure 5. *Comparison of mean item difficulty and mean attribute probability.*



As shown in Figure 5, both the mean item difficulty and mean attribute probability of each cognitive model showed a similar trend. The mean item difficulty of the second attribute [CM1A2: 0.82; CM2A2: 0.71; CM3A2: 0.69] of each cognitive model was lower than that of the first attribute [CM1A1: 0.93; CM2A2: 0.92; CM3A2: 0.91] in each cognitive model. Likewise, the mean attribute probability of the second attribute [CM1A2: 0.88; CM2A2: 0.80; CM3A2: 0.78] of each cognitive model was lower than that of the first attribute [CM1A1: 0.95; CM2A2: 0.96; CM3A2: 0.91] in each cognitive model. This indicates that the mean item difficulty and mean attribute probability decrease as the attribute gets more complex. Besides, this also reflects the linear-shaped cognitive models' hierarchical structure (Alves, 2012) as portrayed in Figure 6.

Figure 6. *Linear hierarchical structure of cognitive model based on the students' responses.*



As shown in Figure 6, the attributes with lower complexity (i.e., CM1A1, CM2A1, and CM3A1) are positioned at the lower hierarchy in the corresponding cognitive models. Meanwhile, the attributes with higher complexity (i.e., CM1A2, CM2A2, dan CM3A2) are positioned at the higher hierarchy in the corresponding cognitive models. Since the two attributes in each cognitive models are related to each other, they are linked together using a straight line and form a linear hierarchical structure. These linear hierarchical structures are also considered as student-based cognitive models (S-CM).

3.2. Dependency among the Attributes in the Cognitive Models

To verify the attribute dependency in each cognitive model, a correlation analysis of the attribute probability for each attribute pair in the cognitive model was conducted. Since the attributes CM1A1, CM2A1, and CM3A1 were negatively skewed ($\text{Skewness}_{\text{CM1A1}} = -2.65$; $\text{Skewness}_{\text{CM2A1}} = -4.78$; $\text{Skewness}_{\text{CM3A1}} = -2.89$), the correlational relationship between each pair of attributes in the six cognitive models were analysed by using Spearman Rank Correlation Coefficient. As illustrated in Figure 6, the correlation coefficients ranged from .97 to .98 indicating that the attribute pair in each cognitive model exhibited a strong positive correlational relationship at the significant level of .05 (Pallant, 2016). This implies that there exists a dependency relationship between the attribute pair in each cognitive model. Hence, the two attributes in each of the cognitive models were positioned next to each other in the attribute hierarchy.

3.3. Consistency between Student-Based and Expert-Based Cognitive Models

The overall consistency between S-CM and E-CM was evaluated based on the HCI computed. Instead of the mean HCI as suggested by Alves (2012), the median HCI was used to represent the heavily left-skewed HCI distributions of the cognitive models constructed in the study with the skewness coefficient ranging from -1.98 to -2.31. The median of HCI for each cognitive model was reported in Table 8. With the median of HCI surpassing the cut score of .80, the six cognitive models derived in this study exhibited excellent fit (Cui et al., 2016). This indicates that the S-CM were highly consistent with the E-CM.

Table 8. *Overall consistency between student-based and expert-based cognitive models.*

Cognitive Model	Skewness	*Md*	Interquartile Range	Interpretation
Cognitive Model 1	-2.08	1.00	(0.64, 1.00)	Excellent
Cognitive Model 2	-1.98	1.00	(0.64, 1.00)	Excellent
Cognitive Model 3	-2.31	1.00	(0.75, 1.00)	Excellent

4. DISCUSSION and CONCLUSION

4.1. To What Extent are Hierarchical Arrangement of Attributes in Expert-Based Cognitive Models Supported by Students' Responses?

The findings indicate that the cognitive models' hierarchical structure developed from the expert perspective was affirmed based on the decreasing trend of mean item difficulty and mean attribute probability which was computed based on the students' responses. This is expected because the attributes were arranged by the experts in increasing order of complexity following the claim made by Iuculano et al. (2018) whereby the mathematics skills are acquired following hierarchical sequences. Since the more basic pre-requisite skill serves as the foundation for mastering a new skill (Iuculano et al., 2018). The new skill is relatively complex. With the increasing attribute complexity, attribute acquisition becomes tougher, and the items become more difficult for the students (Morrison & Embretson, 2014). Thus, the probability of mastering the attributes would be decreased and fewer students would be able to answer the related items correctly. The decreases in mean item difficulty and mean attribute probabilities with the increase of attribute complexity further confirmed the structure of the linearly ordered attribute hierarchy as specified in the E-CMs illustrated in Table 4.

4.2. To What Extent are the Attribute Dependency in The Expert-Based Cognitive Models Supported by Students' Responses?

Meanwhile, the correlational analysis of the attribute probabilities provided convergence evidence to support the hierarchical structure of each cognitive model. This finding is expected because the pre-requisite relationships between the two attributes were exhibited between the two attributes (Sia, 2017). For each cognitive model, the students who must master the second attribute that is more complex are more likely to master the first attribute which is more basic. Since the attribute pair in all cognitive models exhibited a strong positive correlation, we verified the cognitive models' hierarchical structure with the two attributes being positioned adjacent to each other (Sia, 2017).

4.3. To What Extent are the Expert-Based Cognitive Models Consistent with The Student-Based Cognitive Models For 'Subtraction of Time'?

The findings also reveal a high consistency between the S-CMs and E-CMs. This could be due to the use of the curriculum standards and the textbook as the main resources for guiding the attribute specification (Sia et al., 2019). The process of performing the subtraction involving decades and years as well as centuries and years captured in the task analysis is almost similar to the steps shown in the textbook which serve as the main reference in mathematics teaching and learning in the classroom. This could explain the reason underlying the excellent fit among students' responses and the experts' predictions.

4.4. Implications of Study

A total of three cognitive models for 'Subtraction of Time' has been developed in this study. While the attributes' pre-requisite relationship is illustrated in cognitive models, the cognitive models constructed in this study would suggest the instructional sequence for 'Subtraction of time' involving centuries and years' which could foster both conceptual understanding and procedural fluency.

Based on the cognitive models developed, the students should be exposed to the subtraction without regrouping involving measurement of time with the composite unit (i.e., *centuries* and *years*) which serve as an extension from the base-ten subtraction learned in the early grade. Then, the teachers should help the students to recall the relationship between the units of time. Moreover, the underlying reasoning behind regrouping should be explained explicitly when introducing the subtraction of time involving regrouping. With a stronger conceptual understanding, students would have a better procedural fluency in subtraction of time involving *centuries* and *years*.

Besides proposing the instructional sequence, the cognitive models constructed also could be valid to guide the assessment development and make diagnostic inferences related to students' performance on 'Subtraction of Time' involving *centuries* and *years*. This informative diagnostic data would eventually support teachers in the remedial intervention planning for helping the students in overcoming their cognitive weaknesses and thereby foster their mastery of performing subtractions involving these time units.

4.5. Conclusion

To illustrate the cognitive process in performing subtraction of time involving *centuries* and *years*, a total of three cognitive models have been developed in this study. The findings of this study warrant the quality of the cognitive models developed with highly convincing validity evidence which are. Perhaps the findings would encourage the use of cognitive models in guiding the instructional sequence, assessment development and students' result interpretation to support the mathematics teaching and learning for 'Subtraction of Time' involving *centuries* and *years*.

4.6. Limitations and Recommendations

This study is subject to some limitations. Because of the practical constraint, the sample was selected using convenience sampling. Thus, the generalisability of the findings could be reduced. Besides that, the small sample size restricted the choice of the psychometric model used to measure the item difficulty. This eventually reduces the robustness of the findings. To address this limitation, the probabilistic sampling technique is recommended to be used for selecting the larger samples in future studies so that a more robust psychometric model can be applied to calibrate the item difficulty, and the findings would be more generalisable.

Acknowledgments

This study was made possible with funding from the Research University Grant (RUI) Scheme 1001/PGURU/8011027 of Universiti Sains Malaysia, Penang, Malaysia. The authors would like to thank all the teachers and students who voluntarily participated in this study.

Declaration of Conflicting Interests and Ethics

The authors declare no conflict of interest. This research study complies with research publishing ethics. The scientific and legal responsibility for manuscripts published in IJATE belongs to the authors. **Ethics Committee Number**: Human Research Ethics Committee of Universiti Sains Malaysia, USM/JePEM/18030175.

Authorship Contribution Statement

Huan Chin: Conceptualization, Methodology, Data curation, Formal analysis, Visualization, Writing - original draft. **Cheng Meng Chew**: Conceptualization, Methodology, Formal analysis, Writing - review & editing, Supervision, Project administration, Funding acquisition.

Orcid

Huan Chin https://orcid.org/0000-0003-0991-7299
Cheng Meng Chew https://orcid.org/0000-0001-6533-8406

REFERENCES

Akbay, L., Terzi, R., Kaplan, M., & Karaaslan, K.G. (2018). Expert-based attribute identification and validation: An application of cognitively diagnostic assessment. *Journal on Mathematics Education*, *9*, 103-120.

Alves, C.B. (2012). *Making diagnostic inferences about student performance on the Alberta education diagnostic mathematics project: An application of the Attribute Hierarchy Method.* (Publication No. 919011661) [Doctoral Thesis, University of Alberta, Ann Arbor, Canada]. ProQuest Dissertations and Theses database.

Australian Curriculum Assessment and Reporting Authority [ACARA]. (2017). *Numeracy learning progression and history*. https://www.australiancurriculum.edu.au/media/3666/numeracy-history.pdf

Brace, N., Doran, C., Pembery, J., Fitzpatrick, E., & Herman, R. (2019). Assessing time knowledge in children aged 10 to 11 years. *International Journal of Assessment Tools in Education*, *6*(4), 580-591.

Briggs, D.C., & Kizil, R.C. (2017). Challenges to the use of artificial neural networks for diagnostic classifications with student test data. *International Journal of Testing*, *17*(4), 302-321.

Carpenter, T.P., Fennema, E., Franke, M.L., Levi, L., & Empson, S.B. (1999). *Children's mathematics: Cognitively guided instruction.* Heinemann.

Chan, Y.L. (2017). *Super skills modul aktiviti integrasi: Mathematics Year 5 KSSR.* Sasbadi.

Chan, Y.L., Maun, R., & Krishnan, G. (2017). *Dual language programme mathematics Year 5 textbook.* Dewan Bahasa dan Pustaka.

Chen, F., Yan, Y., & Xin, T. (2017). Developing a learning progression for number sense based on the rule space model in China. *Educational Psychology, 37*(2), 128-144.

Chin, H., Chew, C.M., & Lim, H.L. (2021). Development and validation of online cognitive diagnostic assessment with ordered multiple-choice items for 'Multiplication of Time'. *Journal of Computers in Education*, *8*(2), 289-316.

Chin, H., Chew, C.M., & Lim, H.L. (2021b). Development and validation of online cognitive diagnostic assessment with ordered multiple-choice items for 'Multiplication of Time'. *Journal of Computers in Education*, *8*(2), 289-316.

Chin, H., Chew, C.M., Lim, H.L., & Thien, L.M. (2022). Development and validation of a cognitive diagnostic assessment with ordered multiple-choice items for addition of time. *International Journal of Science and Mathematics Education*, *20*(4), 817-837.

Clements, D.H., Sarama, J., Baroody, A.J., & Joswick, C. (2020). Efficacy of a learning trajectory approach compared to a teach-to-target approach for addition and subtraction. *ZDM Mathematics Education*, *52*, 637–648.

Cui, Y., & Leighton, J.P. (2009). The Hierarchy Consistency Index: Evaluating person fit for cognitive diagnostic assessment. *Journal of Educational Measurement*, *46*(4), 429-449.

Cui, Y., Gierl, M., & Guo, Q. (2016). Statistical classification for cognitive diagnostic assessment: An artificial neural network approach. *Educational Psychology*, *36*(6), 1065-1082.

Earnest, D. (2015). When "half an hour" is not "thirty minutes": Elementary students solving elapsed time problem. In T.G. Bartell, K.N. Bieda, R.T. Putnam, K. Bradfield, & H. Dominguez (Eds.), *Proceedings of the 37th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 285-291). Michigan State University.

Earnest, D. (2021). About time: Syntactically-guided reasoning with analog and digital clocks. *Mathematical Thinking and Learning*. [Advance Online Publication].

Fuson, K.C. (1990). Conceptual structures for multiunit numbers: Implications for learning and teaching multidigit addition, subtraction, and place value. *Cognition and Instruction*, *7*(4), 343-403.

Gierl, M.J., Alves, C., & Taylor-Majeau, R. (2010). Using the Attribute Hierarchy Method to make diagnostic inferences about examinees' knowledge and skills in mathematics: An operational implementation of cognitive diagnostic assessment. *International Journal of Testing*, *10*(4), 318-341.

Gierl, M.J., Leighton, J.P., & Hunka, S.M. (2000). An NCME instructional module on exploring the logic of Tatsuoka's Rule-Space Model for test development and analysis. *Educational Measurement: Issues and Practice*, *19*(3), 34-44.

Gierl, M.J., Leighton, J.P., Wang, C., Zhou, J., Gokiert, R., & Tan, A. (2009a). *Validating cognitive models of task performance in algebra on the SAT* (College Board Research 2009-3). The College Board.

Gierl, M.J., Roberts, M.P.R., Alves, C., & Gotzmann, A. (April, 2009b). *Using judgments from content specialists to develop cognitive models for diagnostic assessments*. Paper presented at the Annual Meeting of National Council on Measurement in Education, San Diego, CA.

Gierl, M.J., Wang, C., & Zhou, J. (2008). Using the Attribute Hierarchy Method to make diagnostic inferences about examinees' cognitive skills in algebra on the SAT. *Journal of Technology, Learning, and Assessment*, *6*(6), 1-49.

Gorin, J.S. (2006). Test design with cognition in mind. *Educational Measurement: Issues and Practice*, 25(4), 21-35.

Graf, E.A., Peters, S., Fife, J.H., Van Rijn, P.W., Arieli-Attali, M., & Marquez, E. (2019). *A Preliminary Validity Evaluation of a Learning Progression for the Concept of Function* (Report No: ETS RR–19-21). Wiley.

Hadenfeldt, J.C., Neumann, K., Bernholt, S., Liu, X., & Parchmann, I. (2016). Students' progression in understanding the matter concept. *Journal of Research in Science Teaching*, *53*(5), 683-708.

Harris, S. (2008). It's about time: Difficulties in developing time concepts. *Australian Primary Mathematics Classroom*, *13*(1), 28-31.

Iuculano, T., Padmanabhan, A., & Menon, V. (2018). Systems neuroscience of mathematical cognition and learning: Basic organization and neural sources of heterogeneity in typical and atypical development. In A. Henik & W. Fias (Eds.), *Heterogeneity of function in numerical cognition* (pp. 287-336). Academic Press.

Jin, H., Shin, H.J., Hokayem, H., Qureshi, F., & Jenkins, T. (2019). Secondary students' understanding of ecosystems: A learning progression approach. *International Journal of Science and Mathematics Education*, *17*(2), 217-235.

Kamii, C., & Russell, K.A. (2012). Elapsed time: Why is it so difficult to teach? *Journal for Research in Mathematics Education*, *43*(3), 296-315.

Kane, M.T., & Bejar, I. I. (2014). Cognitive frameworks for assessment, teaching, and learning: A validity perspective. *Psicología Educativa*, *20*(2), 117-123.

Keehner, M., Gorin, J.S., Feng, G., & Katz, I.R. (2017). Developing and validating cognitive models in assessment. In A. Rupp & J.P. Leighton (Eds.), *The handbook of cognition and assessment: Frameworks, methodologies, and applications* (1st ed., pp. 75-101). Wiley Blackwell.

Ketterlin-Geller, L.R., & Yovanoff, P. (2009). Diagnostic assessments in mathematics to support instructional decision making. *Practical Assessment, Research & Evaluation*, *14*(16), 1-11.

Lambert, K., Wortha, S.M., & Moeller, K. (2020). Time reading in middle and secondary school students: The influence of basic-numerical abilities. *The Journal of Genetic Psychology*, *181*(4), 255-277.

Langenfeld, T., Thomas, J., Zhu, R., & Morris, C.A. (2020). Integrating Multiple Sources of Validity Evidence for an Assessment-Based Cognitive Model. *Journal of Educational Measurement*, *57*(2), 159-184.

Leighton, J.P., & Gierl, M.J. (2007). Defining and evaluating models of cognition used in educational measurement to make inferences about examinees' thinking processes. *Educational Measurement: Issues and Practice*, *26*(2), 3-16.

Leighton, J.P., Cui, Y., & Cor, M.K. (2009). Testing expert-based and student-based cognitive models: An application of the Attribute Hierarchy Method and Hierarchy Consistency Index. *Applied Measurement in Education*, *22*(3), 229-254.

Leighton, J.P., Gierl, M.J., & Hunka, S.M. (2004). The Attribute Hierarchy Method for cognitive assessment: A variation on Tatsuoka's Rule-Space Approach. *Journal of Educational Measurement*, *41*(3), 205-237.

Levin, I. (1989). *Principles underlying time measurement: The development of children's constraints on counting time*. In I. Levin and D. Zakay (Eds.), Advances in psychology (Vol. 59, pp. 145-183). Elsevier.

Li, H., & Suen, H.K. (2013). Constructing and validating a Q-matrix for cognitive diagnostic analyses of a reading test. *Educational Assessment*, *18*(1), 1-25.

Linacre, J. (1994). Sample size and item calibration stability. *Rasch Measurement Transactions*, *7*(4), 328.

Morrison, K.M., & Embretson, S.E. (2014). Using cognitive complexity to measure the psychometric properties of mathematics assessment items. *Multivariate Behavioral Research*, *49*(3), 292-293.

Multon, K.D., & Coleman, J.S.M. (2010). *Coefficient alpha*. In N. Salkind (Ed), *Encyclopedia of research design* (pp. 159–162). Sage Publication.

Murata, A., & Kattubadi, S. (2012). Grade 3 students' mathematization through modeling: Situation models and solution models with mutli-digit subtraction problem solving. *The Journal of Mathematical Behavior*, *31*(1), 15-28.

National Council of Teachers of Mathematics [NCTM] (2000*). Principles and standards for school mathematics.* NCTM.

Nichols, P.D. (1994). A framework for developing cognitively diagnostic assessments. *Review of Educational Research*, 64(4), 575-603.

Nichols, P.D., Kobrin, J.L., Lai, E., & Koepfler, J.D. (2017). *The role of theories of learning and cognition in assessment design and development*. In A.A. Rupp & J.P. Leighton (Eds.), The handbook of cognition and assessment: Frameworks, methodologies, and applications (1st ed., pp. 41–74). Wiley Blackwell.

Nuerk, H.C., Moeller, K., & Willmes, K. (2015). *Multi-digit number processing: Overview, conceptual clarifications, and language influences*. Oxford University Press.

Ojose, B. (2015). *Common misconceptions in mathematics: Strategies to correct them*. University Press of America.

Pallant, J. (2016). *SPSS survival manual: A step by step guide to data analysis using SPSS program* (6th ed.). McGraw-Hill Education.

Pellegrino, J.W., & Chudowsky, N. (2003). Focus article: The foundations of assessment. *Measurement: Interdisciplinary Research and Perspectives*, *1*(2), 103-148.

Pellegrino, J.W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment.* National Academy Press.

Pelton, T., Milford, T., & Pelton, L.F. (2018). Developing Mastery of Time Concepts by Integrating Lessons and Apps. In N. Calder, K. Larkin & N. Sinclair (Eds.), *Using Mobile Technologies in the Teaching and Learning of Mathematics* (pp. 153-166). Springer.

Polit, D.F., & Beck, C.T. (2006). The Content Validity Index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing and Health*, *29*(5), 489-497.

Rittle-Johnson, B., & Schneider, M. (2015). Developing conceptual and procedural knowledge in mathematics. In R. Cohen Kadosh & A. Dowker (Eds.), *Oxford handbook of numerical*

cognition (pp. 1102-1118). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199642342.013.014
Russell, M., & Masters, J. (2009, April 13-17). *Formative Diagnostic Assessment in Algebra and Geometry*. Paper presented at the Annual Meeting of the American Education Research Association, San Diego, CA.
Salkind, N. (2010) *Convenience sampling*. In N. Salkind (Ed.), Encyclopedia of research design (p. 254). Sage publications.
Schultz, M., Lawrie, G.A., Bailey, C.H., Bedford, S.B., Dargaville, T.R., O'Brien, G., ... & Wright, A.H. (2017). Evaluation of diagnostic tools that tertiary teachers can apply to profile their students' conceptions. *International Journal of Science Education*, *39*(5), 565-586.
Sia, C.J.L. (2017). *Development and validation of Cognitive Diagnostic Assessment (CDA) for primary mathematics learning of time* [Unpublished master's thesis]. Universiti Sains.
Sia, C.J.L., & Lim, C.S. (2018). *Cognitive diagnostic assessment: An alternative mode of assessment for learning*. In D.R. Thompson, M. Burton, A. Cusi, & D. Wright (Eds.), Classroom assessment in mathematics (pp. 123-137). Springer.
Sia, C.J.L., Lim, C.S., Chew, C.M., & Kor, L.K. (2019). Expert-based cognitive model and student-based cognitive model in the learning of "Time": Match or mismatch? *International Journal of Science and Mathematics Education*, *17*(6), 1–19.
Tan, P.L., Kor, L.K. & Lim, C.S. (2019). Abstracting common errors in the learning of time intervals via cognitive diagnostic assessment. *Creative Practices in Language Learning and Teaching (CPLT) Special Issue: Generating New Knowledge through Best Practices in Computing and Mathematical Sciences*, *7*(1), 3-10.
Tan, P.L., Lim, C.S., & Kor, L.K. (2017). Diagnosing primary pupils' learning of the concept of" after" in the topic" time" through knowledge states by using cognitive diagnostic assessment. *Malaysian Journal of Learning and Instruction*, *14*(2), 145-175.
Tatsuoka, K.K. (1986). Toward an integration of Item-Response Theory and cognitive error diagnosis. In N. Frederiksen, R. Glaser, A. Lesgold, & M.G. Shafto (Eds.), *Diagnostic monitoring of skill and knowledge acquisition* (pp. 453-488). Lawrence Erlbaum Associates.
Tatsuoka, K.K. (1991). *Boolean algebra applied to determination of universal set of knowledge states* (Research Report No: RR-91-44-0NR). Educational Testing Service.
Tatsuoka, K.K. (2009). *Cognitive assessment: An introduction to the Rule Space Method*. Routledge.
Van de Walle, J.A., Karp, K.S., & Bay-Williams, J.M. (2012). *Elementary and secondary school mathematics: Teaching with developmental approach.* Pearson.
Van Steenbrugge, H., Valcke, M., & Desoete, A. (2010). Mathematics learning difficulties in primary education: Teachers' professional knowledge and the use of commercially available learning packages. *Educational Studies*, *36*(1), 59-71.
Wang, C., & Gierl, M.J. (2011). Using the Attribute Hierarchy Method to make diagnostic inferences about examinees' cognitive skills in critical reading. *Journal of Educational Measurement*, *48*(2), 165-187.
Wurpts, I.C., & Geiser, C. (2014). Is adding more indicators to a latent class analysis beneficial or detrimental? Results of a Monte-Carlo study. *Frontiers in Psychology*, *5*(920), 1-15.