ISSN 2087-8885
E-ISSN 2407-0610

Journal on Mathematics Education
Volume 9, No. 1, January 2018, pp. 103-120

EXPERT-BASED ATTRIBUTE IDENTIFICATION AND VALIDATION ON FRACTION SUBTRACTION: A COGNITIVELY DIAGNOSTIC ASSESSMENT APPLICATION

Lokman Akbay[1], Ragip Terzi[2], Mehmet Kaplan[3], Katibe Gizem Karaaslan[1]

[1]Mehmet Akif Ersoy University, Burdur, Turkey
[2]Rutgers, The State University of New Jersey, New Brusnwick, NJ - USA
[3]Artvin Coruh University, Artvin, Turkey
Email: lokmanakbay@gmail.com

Abstract

In this study, we describe the methodology used to identify and validate a set of expert-defined fraction subtraction related attributes. These attributes are expected to be mastered by 6th grade students toward proficiency in fraction subtraction. This research argues and demonstrates that state standards guiding subject instruction plays an important role in the identification of the domain related fundamental attributes. This study also illustrates complete implementation of cognitive diagnosis model framework, which is used to extract diagnostic information about students' specific strengths and weaknesses.

Keywords: Attribute Identification, Cognitive Diagnosis, Diagnostic Classification, Fraction Subtraction

Abstrak

Dalam penelitian ini, kami mendeskripsikan metodologi yang digunakan untuk mengidentifikasi dan memvalidasi sekumpulan operasi pengurangan pecahan yang didefinisikan oleh ahli. Atribut ini diharapkan bisa dikuasai oleh siswa kelas 6, sehingga mereka menguasai pembelajaran pecahan. Penelitian ini menunjukkan bahwa standar negara membimbing pengajaran subjek memainkan peran penting dalam identifikasi atribut dasar terkait domain. Studi ini juga menggambarkan implementasi lengkap kerangka model diagnosis kognitif, yang digunakan untuk mengekstrak informasi diagnostik tentang kekuatan dan kelemahan siswa.

Kata kunci: Identifikasi Atribut, Diagnosis Kognitif, Klasifikasi Diagnostik, Pengurangan Pecahan

How to Cite: Akbay, L., Terzi, R., Kaplan, M., & Karaaslan, K.G. (2018). Expert-based attribute identification and validation: A cognitively diagnostic assessment application. *Journal on Mathematics Education, 9*(1), 103-120.

It is highlighted in the psychometric literature that conventional educational assessments aiming to locate students along a single proficiency continuum do not provide sufficient diagnostic information (de la Torre, 2009; de la Torre, Hong, & Deng, 2010; de la Torre & Karalitz, 2009; Leighton & Gierl, 2007). Useful type of diagnostic information is supposed to help teachers to determine where examinees fail to complete content specific tasks. Educational specialists need formative assessments providing diagnostic information on students' skill mastery to modify classroom instruction and learning (de la Torre & Minchen, 2014). This type of formative assessment is referred to as cognitively diagnostic assessments (CDAs; de la Torre & Minchen, 2014). CDAs yield "interpretative, diagnostic, highly informative, and potentially prescriptive" information (Pellegrino, Baxter, & Glaser, 1999, p.335). Statistical models used in CDA to extract diagnostic information from students' response data are referred to as cognitive diagnosis models (CDMs; de la Torre & Minchen, 20014). These models provide information in the form of examinees' skill profiles. These skill profiles indicate the skills examinees have mastered or not mastered. When skill profiles of examinees are known, remedial actions can be taken toward the skill(s) that are not yet mastered. Increased popularity of CDA gave rise to various approaches, which combine cognitive theory and psychometric practice (e.g., the rule space methodology [Tatsuoka, 1983], the attribute hierarchy method [Leighton, Gierl, & Hunka, 2004], and the generalized deterministic input, noisy "and" gate model framework [de la Torre, 2011]).

One of the earliest attempts for extracting diagnostic information from examinees' response data is in the domain of fraction subtraction. In her research, Tatsuoka (1984, 1990) developed a 20-item test and administered it to 536 middle school students. She used the rule-space methodology (see Tatsuoka, 1983) to identify typical misconceptions in fraction subtraction. In the rule-space methodology, an examinee's observed response pattern is assigned to one of the predetermined knowledge states (i.e., skill mastery profiles), using a particular pattern-recognition method. To accomplish this, Tatsuoka identified the following skills needed to solve each fraction subtraction item:

(S1) convert a whole number to a fraction,

(S2) separate a whole number from fraction,

(S3) simplify before subtracting,

(S4) find a common denominator,

(S5) borrow from whole part,

(S6) column borrow to subtract the second numerator from the first,

(S7) subtract numerators,

(S8) reduce answer to simplest form.

For instance, an item (3/4–3/8=?) required two specific skills to answer it correctly: S4: find a common denominator (3/4–3/8=6/8–3/8) and S7: subtract numerators ((6-3)/8=3/8). More information on how CDA provides information on students' strengths and weaknesses is given in the next section.

The domain of fraction subtraction, or in general the fractions, is of the essence in mathematics as it is a substantial complement of algebra. The mathematics literature emphasizes unignorable relationships between the fractions and some other mathematical domains, including ratio and proportion, proportional reasoning, decimal numbers, and measurement (Haser & Ubuz, 2002; Walle, Karp, & Bay-Williams, 2013). Furthermore, Ni (1999) argued that comprehension of fractions could have positive impact on the development of problem solving skills and understanding of more complex and advanced topics in algebra. As it is crucial, researchers claim that fractions and operations in fractions are among the most difficult topics in elementary and secondary educational contents (Behr, Harel, Post, & Lesh, 1992; Misquitta, 2011; Son, 2012; Son & Crespo, 2009; Tirosh, 2000). Over-generalization of the mathematical operation routines is among the common research results explaining why fraction related problem solving is challenging (Haser & Ubuz, 2002; Mack, 1995; Vinner, Hershkowits, & Bruckheimer, 1981). Furthermore, failure in coexistence of procedural and conceptual knowledge may also be counted among the main reasons for having difficulties in completing fraction tasks (Birgin & Gürbüz, 2009).

Given the importance of and difficulties in learning fractions, identifying students' strengths and weaknesses through formative assessment is important for adjusting teaching and learning activities. The skills needed to complete fraction subtraction tasks have been defined earlier by Tatsuoka (1984, 1990). However, some skills and strategies may change in conjunction with an alternative curriculum. As such, we argue that the skills defined based on a particular state-standards (or curricula) may not be useful for obtaining diagnostic information when borrowed as it is. Further, the skills and strategies needed for proficiency may vary by grade level. Therefore, by collaborating with domain experts in this research, we aim to identify and validate the necessary skills for deeming a student proficient in the fraction subtraction domain in Turkey. We expect that these skills can serve as a foundation for designing effective measurement tools in diagnosing what skills or pieces of knowledge are required for six-graders to master in Turkish fraction subtraction curriculum. To meet our expectation, we are interested in both identifying and validating a skill-set. We conclude the study by developing a cognitively diagnostic test that could be useful to promote student learning beyond just measuring it.

Overall, this study primarily discusses the process of identification and validation of six-grade fraction subtraction skills by expanding on the methodology. Then, we evaluate the cognitive model through a provisional test where a set of response data are analyzed with deterministic input, noisy “and” gate (DINA; de la Torre, 2009; Junker & Sijtsma, 2001) model. On top of developing a cognitive model by identifying and validating necessary skill-set, this research is of value as it displays all necessary steps that must be taken throughout a CDA application. Lastly, this study further shows how two skill-sets can differ by state standards that guide school subject instruction.

The CDA is conducted for the purpose of improving and directing teaching and learning activities (DiBello & Stout, 2007) in accordance with the crucial feedback obtained from the analysis of assessment results. Several specific and general cognitive diagnosis models (CDMs) merging cognitive theory and testing design were proposed. In DMC literature, a generic term “attribute” is used to refer to the cognitive processes, skills, knowledge representations, and problem solving steps (de la Torre, 2009; de la Torre & Lee, 2010) measured by CDA. Accordingly, in this manuscript, *attribute* will be used to refer to the skills, abilities, and cognitively conceptual or operational knowledge representations that examinees are required to master to successfully complete domain specific tasks.

A common component of these CDMs is an item-by-attribute specification matrix or the Q-matrix (Tatsuoka, 1985, 1990). More specifically, rows of this matrix correspond to the items and indicate the required attributes to successfully complete the corresponding item. For attributes $k = 1,\ldots, K$ and items $j = 1,\ldots, J$ in a test, the Q-matrix element q_{jk} is specified as

$$q_{jk} = \begin{Bmatrix} 1, \text{if item j requires attribute k} \\ 0, \text{otherwise} \end{Bmatrix}.$$

Remember that eight attributes were defined in total for fraction subtraction. The q-vector for the item in the earlier example (3/4–3/8=?) can be written as 00010010. This vector specifically shows that this item requires the fourth and seventh attributes for a correct response.

Using the observed responses given to the test items, CDMs assign a vector of attribute mastery scores to each examinee. These attribute mastery scores are typically binary and indicate whether an examinee possesses corresponding attributes. More generally, based on the attribute mastery or no mastery status, a test measuring K attributes defines 2^K different attribute mastery profiles. Using an appropriate CDM, examinees' attribute profiles can be estimated by classifying examinees into one of the 2^K latent classes defined by K attributes. These profile estimates indicate which particular attributes have or have not been mastered by examinees. Therefore, rather than having a single scale score in general ability (e.g., a math score of 80 out of 100), we can have information on specific attributes an examinee has or has not mastered (e.g., you are able to find a common denominator but you do not know how to convert a whole number to fraction).

For instance, when three attributes are measured in a test, an item can require either a single attribute or multiple of them. Similarly, for three attributes, examinees must have one of the eight (i.e., 2^3) distinct latent profiles (i.e., {000}, {100}, {010}, {001}, {110}, {101}, {011}, and {111}). By employing an appropriate CDM, each examinee is assigned with an attribute profile estimate that corresponds to the latent class the examinee is classified to. For example, an examinee classified to the first latent class will have the first attribute profile (i.e., {000}), which can be interpreted such that this examinee does not possess any of the measured attributes. Similarly, attribute mastery profile of {100} is interpreted such that examinees possess the first but not the second and the third attributes. The remaining attribute profiles are interpreted in the same vein.

Although the types and psychometric properties of CDMs are not in the scope of this paper, it should be noted that various specific and general CDMs have been recently proposed. A comprehensive discussion on CDMs can be found in Rupp and Templin (2008) and de la Torre (2011). Because it is used in the attribute validation process of this study, a specific model referred to as the DINA model (de la Torre, 2009; Junker & Sijtsma, 2001) is briefly explained below.

The DINA Model

The DINA model is known as one of the most parsimonious and interpretable cognitive diagnosis models. This model has only two item parameters (i.e., guessing and slip), stand for probability of success when students do not possess all required attributes and probability of incorrect answer when students have mastered all required attributes, respectively. Based on the DINA model response function, examinees missing only one of the several required attributes will have the same probability of correct response as the ones who lack all required attributes (de la Torre, 2009; Rupp & Templin, 2008). This is statistically represented by the conjunctive condensation function (Maris, 1995, 1999). Given an examinee's attribute profile, $\boldsymbol{\alpha}_i$, and the j^{th} row of the Q-matrix (i.e., attribute specification for item-*j*), the conjunctive condensation rule generates a deterministic response (η_{ij}=1 or 0) through the function

$$\eta_{ij} = \prod_{k=1}^{K} \alpha_{ik}^{q_{jk}}. \quad (1)$$

The item response function of the DINA model allows possibility of slip when an examinee has all required attributes, and it allows an examinee to guess when s/he lacks at least one of the required attributes. Slip and guessing for item *j* are denoted as

$$s_j = P(X_{ij} = 0|\eta_{ij} = 1) \text{ and } g_j = P\left(X_{ij} = 1\middle|\eta_{ij} = 0\right),$$

respectively, where X_{ij} is the observed response of examinee *i* to item *j*. Given s_j and g_j, the DINA model item response function is written as

$$P\left(X_{ij} = 1\middle|\boldsymbol{\alpha}_i\right) = P\left(X_{ij} = 1\middle|\eta_{ij}\right) = g_j^{(1-\eta_{ij})}(1 - s_j)^{\eta_{ij}}, \quad (2)$$

where $\boldsymbol{\alpha}_i$ is attribute pattern that is possessed by examinee *i*; η_{ij} indicates the expected response of examinee *i* to item *j*, respectively (de la Torre, 2009).

METHOD

Mathematics knowledge is classified as conceptual and procedural knowledge and students are considered fully-proficient when they acquire both types of highly related mathematical knowledge (Hiebert & Lefevre, 1986). Although direct measurement of the cognitively conceptual knowledge may be challenging, it could be measured through observable (e.g., procedural) knowledge. Therefore, in this research, we are interested in measuring the attributes that may reflect both conceptual and procedural knowledge.

Cognitive process of human performance can be identified and validated in two ways; 1. by conducting verbal data analyses (e.g., interviews and think-aloud protocols for problem solving) and 2. by relying on expert thoughts via content analysis (Leighton, Cui, & Cor, 2009). These two potentially useful methods can be distinguished by their top-down and bottom-up natures (Chi, 1997; Ericsson, 2006; Leighton, 2004). In a top-down approach, literature review can be conducted to find out existing learning theories that may guide researchers in the process of attribute identification (Embretson, 1998; Leighton et al., 2009). In situations where theories are absent or insufficient, review of curriculum objectives and learning outcomes can also be useful for attribute identification (Leighton et al., 2009). In such cases, individuals reviewing the curriculum objectives and learning outcomes need to have teaching experience, which may allow them to have insight into attributes students would use to complete domain related tasks.

A bottom-up approach can also be developed for identification and validation of attributes. In this approach, data about examinee response process are directly collected via interview and think-aloud procedures (Chi, 1997; Leighton et al., 2009). Then, experts and researchers strive for identifying common arguments and themes in the cognitive processes, skills, and knowledge that may be used to complete specific tasks. Note that the top-down approach is theoretically based, whereas the bottom-up approach is data-driven. In this study, we have adopted the top-down approach for attribute identification and the bottom-up approach for attribute validation.

The methodology of this study consists of three parts. The first part is devoted to identification of pedagogically meaningful and psychometrically measurable attributes in the domain. The goal of this part is to explore the conceptual and/or procedural attributes six-grade students need to master toward proficiency. It should be noted here that besides the authors of the manuscript who have background in mathematics, teaching, and psychometrics; four elementary mathematics teachers who are working in public schools in Turkey and a mathematics education specialist were involved in this study. The mathematics education specialist and teachers will be referred to as the expert group from now on. The second and the third parts of the methodology are allotted to validation of the attributes defined in the first part. This is accomplished by employment of a verbal data analysis and evaluation of the cognitive model in the second and third parts, respectively.

More specifically, in the second part, experts solved a set of prototype items providing detailed solution steps to confirm the match between items and attributes. As experts solved items, they also took great care to use the skills and strategies taught in the state schools. Afterwards, these 20 items were randomly grouped into four small tests, each of which had five items. Each of these four tests was administered to two groups of randomly selected six-graders from among eight schools, which were conveniently selected across the country. Group sample sizes varied from N = 6 to N = 10. Therefore, each item was administered to a total of 12 to 20 students. The goal in this test administration was to understand how six-graders perform on fraction subtraction tasks. Therefore, students were asked to write down all the operations and calculations they go through to reach an answer. They were also asked to tell in detail what they think as they solve the items.

We further attempted to collect evidence for unidentified attributes, if any, used in problem solving. It should be noted here that any evidence toward the use of omitted attributes was to motivate researchers to look back and review the attribute-set. Therefore, we can claim that in its entirety, the process of attribute identification and item specification (i.e., matching the items and their required attributes) was an iterative procedure such that attributes and items helped modifying one another.

As the third part of the study, we intended to construct a test to assess the cognitive model specified by the six attributes that are required to perform successfully on fraction subtraction tasks. Thus, the expert group was asked to develop a *provisional* test using the prototype items as well as novel items. To ensure that the test fits the purpose, researchers emphasized some desirable features of this type of CDA tests. Inclusion of all possible single-attribute items along with items with various combinations of the attributes, and measuring each attribute roughly equal number of times are among these features. This test with 29 items was called provisional because the ultimate goal was to refine it by adjusting them based on the preliminary analysis of the data gathered. Then, when administered, the final test will hopefully provide sufficient and accurate diagnostic information. The Q-matrix describing the association between the items and attributes is given in Table 1. For example, item nine;

$$2\frac{2}{5}-\frac{4}{5}=? \qquad (3)$$

requires the fifth (convert a mixed fraction to improper fraction) and sixth (subtract numerators) attributes. Thus, for a correct response, an examinee needs to know how to "convert a mixed fraction to improper fraction" and how to "subtract numerators".

Table 1. The Q-matrix for the provisional test

Item	A1	A2	A3	A4	A5	A6	Item	A1	A2	A3	A4	A5	A6
1	1	0	0	0	0	0	16	0	1	0	0	0	0
2	0	1	0	0	0	0	17	0	0	1	0	1	1
3	0	0	1	0	0	0	18	0	0	0	1	0	1
4	0	0	0	1	0	0	19	1	0	0	0	0	0
5	0	0	0	0	1	0	20	1	1	1	1	0	1
6	0	0	0	0	0	1	21	0	0	0	1	0	0
7	0	0	1	0	0	0	22	0	0	0	0	1	0
8	0	0	1	1	0	1	23	1	0	0	1	0	1
9	0	0	0	0	1	1	24	1	0	0	1	0	1
10	0	0	0	1	0	1	25	1	0	0	1	0	1
11	0	0	0	0	1	1	26	0	1	1	0	1	1
12	1	1	0	1	0	1	27	0	0	1	0	0	1
13	0	0	1	0	0	0	28	0	1	0	0	0	0
14	0	1	1	0	1	1	29	0	0	1	0	0	0
15	0	0	0	0	0	1							

Note. A1 = convert a whole number to fraction; A2 = convert an improper fraction to a mixed fraction; A3 = simplification; A4 = finding a common denominator; A5 = convert a mixed fraction to improper fraction; A6 = subtract numerators.

Although communication between the researchers and the expert group was dynamic and continuous, we held a number of large meetings to discuss and direct different phases of the study. Aim and outcome of each of these extensive meetings were elucidated in the following section.

RESULTS AND DISCUSSION

Part I: Attribute Identification

Meeting I: In the first meeting, researchers and expert group members gathered together to set the objectives and a course for the research project. One of the accomplishments in this meeting was to come up with a relatively small and target-specific reference list for literature review. The attendees finalized the reference list by focusing exclusively on fraction subtraction for elementary mathematics education. The first meeting was followed by the literature review to support and document the reasoning toward attribute identification.

Meeting II: Prior to the second meeting, the final report on literature review was sent to expert group members. Based on the literature review, several candidate attributes were identified as they cut across various types of fraction subtraction problems. For example, the literature claimed that students had difficulties in subtracting the fractions when one or both fractions had a whole part (Mack, 1995). Thus, a whole part in a fraction subtraction item can make difference in student responses. Therefore, "converting a mixed fraction to an improper fraction" became a candidate attribute. The highlights gathered from the textbooks in use are given in Table 2. Note that only points regarding the fraction subtraction are given in this table.

Table 2. Highlights from the textbooks

Fifth Grade Mathematics Textbook (Ministry of National Education - Pasifik Publication)
1. Compare and contrast a natural number with a fraction,
2. A natural number is a fraction with 1 in the denominator,
3. To convert a mixed fraction to improper fraction, whole part is written as fraction and it is added to remaining fraction part,
4. To convert an improper fraction to a mixed fraction, the numerator is divided by the denominator,
5. Dividing or multiplying both the numerator and denominator of a fraction with the same number does not change the value of the fraction,
6. Equivalent fractions are formed by either multiplying or dividing both the numerator and denominator with the same number,
7. Fraction subtraction is performed by subtracting the numerators of the fractions when they have a common denominator.
Sixth Grade Mathematics Textbook (Ministry of National Education - Sevgi Publication)
1. Equalize the denominators before subtracting the numerators of the fractions,
2. After equalizing the denominators of the two fractions, subtraction operation is performed in the same way with equal denominator cases,
3. When subtracting fractions with equal denominators, difference in the numerators becomes the numerator of the resulting fraction and the common denominator is kept.

The researchers and experts agreed upon the need for reviewing the fraction subtraction related learning objectives for five- and six-graders. The subject curriculum is developed and the learning objectives are set by the official commissions established by the Board of Education and Discipline of National Education (Milli Eğitim Bakanlığı Talim ve Terbiye Kurulu). These commissions may consist of subject specialists, teachers, academic faculty members, and other domain experts (Milli Eğitim Bakanlığı Talim ve Terbiye Kurulu Başkanlığı Yönetmeliği, 2012). After the second meeting, expert group has completed their review on the subject curriculum and learning objectives. After that, the third meeting was scheduled to discuss the learning objectives and their relation to the literature.

Meeting III: At this meeting, experts and researchers listed domain related learning objectives set by the commissions established by the board of education and discipline. Among all the learning objectives defined for fractions, we only considered subtraction related items. Table 3 presents our final list of objectives. Moreover, the researchers and expert group decided to review the items in the textbooks and various workbooks to make sure that the candidate attributes conform to existing operational items. Among reviewed items (i.e., approximately 400 items), 50 representative items were selected to be discussed in the fourth meeting. Items in the sample varied in terms of required attribute combinations.

Table 3. Fraction subtraction related learning objectives

Fifth Grade Objectives
"SSBATa understand that a mixed fraction is the sum of a natural number and a proper fraction then convert a mixed fraction to improper fraction and vice versa"
"SSBAT comprehend the fact that the value of a fraction is not affected when both the numerator and denominator multiplied or divided by the same number"
"SSBAT perform and comprehend subtraction in fractions with equal and unequal denominators"
Sixth Grade Objectives
"SSBAT solve problems requiring subtraction in fractions with equal and unequal denominators"
"SSBAT extend a fraction by multiplying the fraction with natural numbers"

SSBAT = students should be able to.

By discussing the literature highlights, learning objectives, and the item types to outline the attribute-set, research participants have composed the following list:

A1. convert a whole number to a fraction,
A2. convert an improper fraction to a mixed fraction,
A3. simplification,
A4. finding a common denominator,
A5. convert a mixed fraction to improper fraction,
A6. subtract numerators.

Furthermore, experts believed that the type of expression of the problem (i.e., verbal expression or mathematical statement) would not really make difference for the fifth and/or sixth grade students as they are expected to conceptualize fractions in earlier stages. In other words, students must have prior knowledge on stating a verbally expressed problem in fractional notations and vice versa.

Part II: Think-aloud Protocol toward Attribute Validation

To make sure which attributes are needed to complete the prototype items, the experts detailed items' solution steps. As experts solved the items, they also took great care to use the skills and strategies taught in the state schools. Afterwards, these 20 items randomly grouped into four small tests. Each of these small tests was administered to two groups of randomly selected six-graders among eight conveniently selected schools. The goal in this test administration was to understand how six-graders perform on fraction subtraction tasks. Therefore, students were asked to write down all the operational steps as they solve problems to reach an answer. They were also asked to tell in detail what they think within each step.

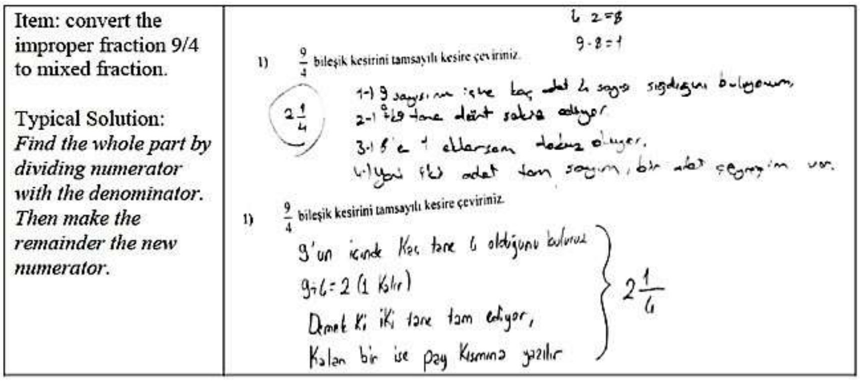

Figure 1. Typical student responses to item 1

The prototype items used in think-aloud protocol targeted either a single or multiple of previously identified six attributes. For example, item 1 was developed to see how six-graders convert an improper fraction (i.e., a fraction with a numerator larger than its denominator) to a mixed fraction (i.e., a whole number with a proper fraction). This item and two exemplars for typical student responses are given in Figure 1. Likewise, items 2, 4, 8, 10, and 11 were developed to measure single attributes (i.e., finding a common denominator, simplification, subtract numerators, convert a whole number to a fraction, and convert a mixed fraction to improper fraction, respectively). Experts also developed more complex prototype items, for which students needed to use multiple attributes. As an example, an item requiring three attributes and typical student responses are given in Figure 2. After test administration, research group scheduled the fourth meeting to discuss student solutions and the attributes they actually used at any steps of the solution process. Thus, experts were given the student solutions prior to fourth meeting to code students' solution steps and corresponding attributes used.

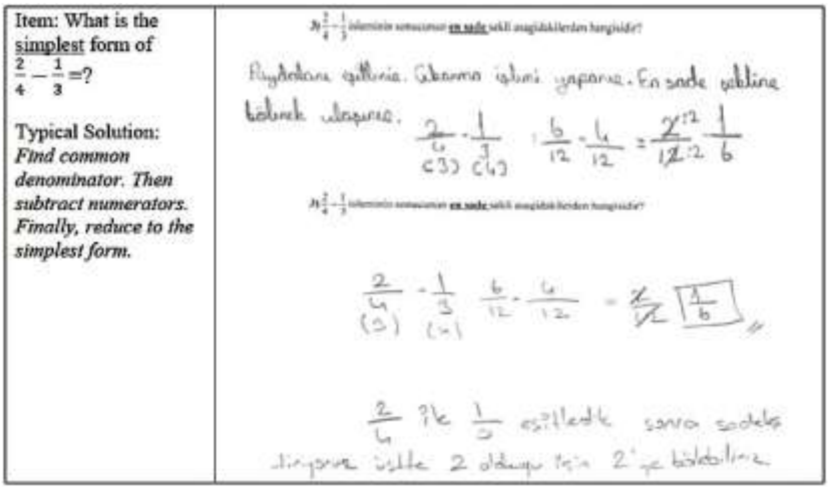

Figure 2. Typical student responses to an item requiring three attributes

Meeting IV: In this meeting, experts compared their coding and discussed the associations between attributes. Then, they compared the observed task-by-attribute associations with a priori specifications (i.e., expert expectations on the attributes required to perform successfully on each task). Experts and researchers agreed that all students who answered these items correctly used the attributes specified by the experts. Therefore, sufficient and convincing evidence were gathered to claim that these identified attributes were necessary and sufficient for six-graders to master toward proficiency in fraction subtraction.

In addition, a considerable amount of students' responses signaled a misconception on converting a whole number to a fraction (i.e., A1). This is shown in Figure 3 where a typical student response to an item that requires converting the whole number 9 to fraction. Although the response data support the claim that there is a misconception in mastering A1; this has nothing to do with the identified attribute and is, in fact, a common problem as it was discussed by Mack (1995). She noticed similar patterns and explained the possible underlying reasons for this type of responses (e.g., student might think that a fraction represents the pieces of a whole, therefore, cannot be larger than 1). At the end of this meeting, researchers fully informed the expert group on the work of Tatsuoka (1984, 1990) and the attributes she identified in her study. Then, experts were asked to compare and contrast the two sets of attributes by the fifth meeting.

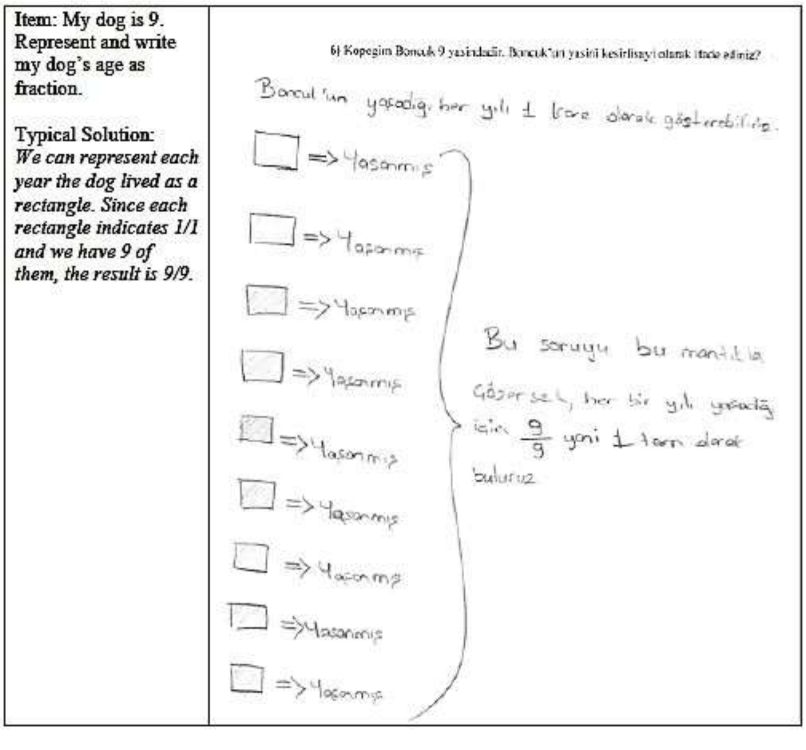

Figure 3. Student response indicating a misconception in mastering A1

Meeting V: Experts and researchers discussed similarities and disparities of the two sets of attributes. They further discussed the viability of Tatsuoka's attributes in Turkish curriculum (i.e., whether the set of attributes defined by Tatsuoka (1984, 1990) represents the skills and knowledge taught and used for elementary fraction subtraction). The following are the highlights from this meeting:

1. Experts agreed that three attributes in Tatsuoka's set (i.e., convert a whole number to fraction, find a common denominator, and subtract numerators) are the same and necessary attributes based on Turkish math curriculum.
2. *Simplify before subtracting* and *reduce answers to simplest form* were merged and referred to as *simplification*. Experts argued that these two attributes eventually stand for dividing numerator and denominator with a common number to ease computations. They also stated that if students know how to simplify and why to simplify, it does not matter at what solution step they simplify to obtain a simpler answer.
3. Similarly, *separate a whole number from fraction*, *borrow from whole number part*, and *column borrow to subtract the second numerator from the first* were replaced by one attribute and called *converting a mixed fraction to improper fraction*. Indeed, once we convert a mixed fraction to improper fraction, there is no longer need for borrowing from whole part or column borrowing to subtract the second numerator from the first.
4. Lastly, our study included a novel attribute referred to as *convert an improper fraction to a mixed fraction*.

Based on the literature review and examination of the textbooks, experts claimed that students are not necessarily taught to separate whole number from fraction. Rather, they are taught to convert a mixed fraction to improper fraction. Therefore, students would not need skills such as *borrowing from whole part* or *column borrows to subtract the second numerator from the first*. Textbooks touch on how subtraction tasks involving mixed fractions might be solved by borrowing from the whole part; however, solutions employing this attribute are usually given as an alternative to the solution that involves converting a mixed fraction to an improper one. Thus, the mentioned attributes (i.e., separate a whole number from fraction, borrow from whole number part, and column borrow to subtract the second numerator from the first) are not necessarily among the essential attributes examinees would use in elementary school level. It should be noted here that students' solutions for prototype items supported these arguments.

After all, the identification and validation procedures of fraction subtraction domain yielded six attributes that are needed to deem someone proficient. These finalized attributes are: (A1) convert a whole number to a fraction, (A2) convert an improper fraction to a mixed fraction, (A3) simplification, (A4) finding a common denominator, (A5) convert a mixed fraction to improper fraction, and (A6) subtract numerators.

Part III: Model Evaluation toward Attribute Validation

A test with 29 items was administered to 255 randomly selected sixth grade students from six conveniently selected state-schools. All items were multiple choice items and the responses were coded as 0 and 1 stating incorrect and correct answers, respectively. The tetra choric correlations among the items were calculated using the version 1.5.8 of R-package psych (Revelle, 2015). The tetra choric correlation matrix is given in Table 4. In the second column of this table, items are tagged such that items requiring the same set of attributes had the same tag. We expected high correlation between the items holding the same tag assuming the Q-matrix was correctly specified and items were sufficiently discriminative. When item is discriminative, examinees who possess the required attribute(s) would give a correct response whereas examinees who do not have the required attribute(s) would fail to provide a correct answer.

Table 4. Tetra choric correlations among the items

Items	T	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
1	A	1.00																												
2	B	.43	1.00																											
3	C	.21	.31	1.00																										
4	D	.31	.46	.14	1.00																									
5	E	.44	.69	.34	.39	1.00																								
6	F	.30	.44	.32	.53	.57	1.00																							
7	C	.24	.44	.31	.21	.43	.25	1.00																						
8	G	.31	.39	.35	.26	.36	.53	.26	1.00																					
9	H	.58	.58	.32	.36	.59	.42	.33	.48	1.00																				
10	I	.49	.76	.31	.43	.65	.42	.55	.47	.70	1.00																			
11	H	.54	.67	.38	.37	.70	.34	.27	.41	.86	.75	1.00																		
12	J	.35	.35	.31	.31	.25	.40	.28	.41	.40	.45	.33	1.00																	
13	C	.35	.38	.03	.41	.50	.32	.21	.14	.33	.46	.44	.32	1.00																
14	K	-.01	.24	.09	.19	.01	.04	.12	.10	-.05	.08	.00	.19	.21	1.00															
15	F	.32	.28	.14	.24	.42	.61	.30	.17	.31	.32	.40	.37	.19	.09	1.00														
16	B	.26	.59	.12	.18	.61	.33	.40	.28	.56	.50	.51	.26	.27	.09	.31	1.00													
17	L	.36	.48	.19	.24	.28	.34	.24	.49	.47	.44	.51	.28	.33	.19	.22	.36	1.00												
18	I	.17	.50	.29	.31	.50	.41	.37	.31	.46	.58	.51	.29	.39	.24	.32	.45	.35	1.00											
19	A	.64	.41	.13	.38	.43	.42	.43	.29	.39	.48	.37	.34	.40	.10	.36	.40	.20	.31	1.00										
20	M	.56	.34	.32	.31	.52	.23	.34	.22	.53	.57	.61	.35	.38	-.01	.20	.43	.31	.31	.49	1.00									
21	D	.42	.51	.24	.60	.53	.53	.18	.46	.39	.52	.50	.38	.33	.06	.41	.37	.26	.46	.45	.38	1.00								
22	E	.56	.65	.40	.35	.79	.44	.31	.37	.68	.71	.76	.33	.45	-.07	.41	.63	.36	.42	.51	.56	.61	1.00							
23	N	.50	.60	.08	.50	.46	.56	.27	.37	.50	.66	.50	.46	.50	.15	.38	.43	.43	.42	.47	.49	.50	.55	1.00						
24	O	.31	.57	.09	.42	.38	.46	.19	.37	.36	.60	.46	.20	.40	.35	.33	.40	.37	.29	.41	.41	.53	.47	.74	1.00					
25	O	.35	.45	.14	.45	.51	.56	.28	.42	.51	.66	.55	.37	.39	.19	.42	.45	.36	.35	.43	.40	.48	.62	.68	.67	1.00				
26	K	.40	.39	.30	.21	.32	.03	.31	.37	.48	.39	.51	.35	.40	.29	-.01	.38	.29	.28	.37	.42	.46	.44	.36	.30	.33	1.00			
27	P	.33	.66	.35	.42	.64	.54	.27	.45	.52	.68	.64	.41	.38	.23	.40	.49	.57	.33	.48	.58	.57	.68	.55	.61	.57	.32	1.00		
28	B	.46	.74	.26	.39	.63	.40	.32	.34	.57	.80	.64	.41	.30	.15	.27	.63	.32	.48	.46	.57	.62	.70	.67	.62	.54	.43	.70	1.00	
29	C	.31	.49	.45	.41	.55	.48	.39	.35	.52	.68	.60	.41	.39	.17	.41	.45	.39	.40	.45	.41	.56	.56	.42	.51	.63	.26	.70	.56	1.00

Note: T = tag that is used to label the items with the same q-vector

With few exceptions, the results were consistent with our expectation. For instance, items 1 and 19 have the same tag (i.e., A), which stands for q-vector 100000, and the correlation between them is .64. Similarly, items 5 and 22 have the same tag (i.e., E with q-vector 000010) and they have a correlation of .79. Furthermore, two items requiring the same multiple attributes are 24 and 25 and are tagged as O. The correlation between these two items is .67. As is seen, we adopted pairwise correlations between the items with the same tag as a criterion to support the claim that these items require the same sets of attributes for success. Although correlation between items with the same tag was high in general, there were some exceptions such as items tagged C, which states that both these items require the third attribute (i.e., simplification). Items 3, 7, 13, and 29 were in this group and the average pairwise correlation among them was about .30. The aberrant item parameters (e.g., too high guessing or slip), item representation, item distractors, and misconceptions might have caused these low correlations. However, this unexpected result requires further investigation.

Based on the DINA model estimation, which is described by de la Torre (2009), we have obtained the item and person parameter estimates. Item parameter estimates are given in Table 5. For many items, the item parameter estimates fell within reasonable ranges, which are $g_j = .25 \pm .15$ and $s_j < .15$ for guessing and slip parameters, respectively.

Table 5. Item parameters with standard errors

Items	g	SE_g	s	SE_s	Items	g	SE_g	s	SE_s
1	0.22	0.04	0.2	0.04	16	0.35	0.05	0.1	0.03
2	0.33	0.05	0.06	0.02	17	0.22	0.03	0.33	0.05
3	0.38	0.04	0.27	0.04	18	0.43	0.04	0.1	0.03
4	0.47	0.06	0.08	0.02	19	0.3	0.04	0.13	0.03
5	0.39	0.05	0.03	0.02	20	0.15	0.03	0.32	0.05
6	0.72	0.04	0.06	0.02	21	0.34	0.06	0.03	0.02
7	0.24	0.04	0.47	0.05	22	0.23	0.05	0.03	0.02
8	0.18	0.03	0.42	0.05	23	0.4	0.04	0.01	0.01
9	0.2	0.04	0.05	0.02	24	0.48	0.04	0	0.09
10	0.27	0.04	0.02	0.01	25	0.42	0.04	0	0.03
11	0.12	0.03	0.02	0.01	26	0.12	0.03	0.56	0.05
12	0.34	0.04	0.21	0.04	27	0.23	0.04	0.09	0.03
13	0.17	0.03	0.49	0.05	28	0.22	0.04	0.05	0.02
14	0.2	0.03	0.74	0.04	29	0.28	0.04	0.05	0.03
15	0.71	0.05	0.05	0.02					

Note: g = guessing parameter; s = slip parameter; SE_g *= standard error for guessing parameter;* SE_s *= standard error for slip parameter.*

Due to the four-option multiple choice nature of the test, examinees who did not master the required attributes for an item had a 1/4 chance of answering any item correctly. Thus, we expected the guessing parameters be around .25, and the slip parameters to be close to its lower bound. However, due to the small sample size and possible measurement errors, a deviation up to .15 for either slip or guessing parameter might still be considered as sufficient evidence to conclude that item requires neither more nor less attributes than the ones specified in the Q-matrix. Furthermore, as it was the case with tetra choric correlations, items 3, 7, and 13 were among the items with very high slip such that some examinees who were supposed to complete these items successfully could not find the correct answers. Consequently, this point should be examined thoroughly.

CONCLUSION

Cognitively diagnostic assessments need to be deliberately developed to determine students' mastery or nonmastery status of cognitive competencies, skills, and strategies. As argued in de la Torre and Minchen (2014), CDM framework can be used to design and develop CDAs that can provide sufficient and accurate diagnostic information. After obtaining the diagnostic information, it can be used to modify classroom instruction and learning. Furthermore, because the diagnostic information would be in the person level, tailored remedial actions can also be taken to help students master the attributes they have not mastered yet. The successful implementation of CDM framework in an operational setting would depend heavily on accuracy and validity of attributes. Therefore, attribute identification and validation procedures in a CDA application cannot be overemphasized. In this respect, we provided an example of expert-based attribute identification and validation procedures to guide practitioners on this matter.

We claimed and manifested that attributes required to complete domain relevant tasks rely heavily on curriculum or state standards based on which students are taught specific skills and strategies. This study demonstrates that, to diagnose students' strengths and weaknesses, we may not always rely on a set of attributes defined in concordance with a particular curriculum in a country. Here, based on the Turkish elementary math curriculum, we have identified and validated six attributes that students need to master for proficiency in fraction subtraction domain. However, Tatsuoka (1984, 1990) claimed that eight attributes were needed to deem someone proficient in this domain. As in this study, attribute sets in other subjects may also have variations due to differences in instructions.

It should be noted here that the grade level of the target group may also have a substantial impact on the identification of attributes. This is particularly true when teaching sequence of attributes are not the same across different state or district standards. Therefore, the grade level is another factor that needs to be taken into account before basing assessment on the readily available attributes-set. Consequently, researchers and practitioners need to be aware that relying on a set of previously identified attributes in cognitive diagnosis modeling may not always provide the best outcomes.

Moreover, given the relative novelty of the CDM framework, this study also aimed to provide a throughout application of the framework that may provide a guideline for practitioners to implement their own assessments in various domains. Lastly, the study is particularly of value as it provides the set of attributes based on which item-by-attribute matrix (i.e., Q-matrix) can be created for CDM analysis to extract diagnostic information on Turkish six-graders' fraction subtraction acquisition.

REFERENCES

Behr, M. J., Harel, G., Post, T., & Lesh, R. (1992). Rational number, ratio and proportion. In D. Grows (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 296-333). New York, NY: Macmillan Publishing Company

Birgin, O., & Gürbüz, R. (2009). İlköğretim II. kademe öğrencilerinin rasyonel sayılar konusundaki işlemsel ve kavramsal bilgi düzeylerinin incelenmesi. *Uludağ Üniversitesi Eğitim Fakültesi Dergisi, 22,* 529-550.

Chi, M. T. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *The Journal of the Learning Sciences, 6,* 271-315.

de la Torre, J. (2009). DINA model and parameter estimation: A didactic. *Journal of Educational and Behavioral Statistics, 34,* 115-130.

de la Torre, J. (2011). The generalized DINA model framework. *Psychometrika, 76,* 179-199.

de la Torre, J., & Karelitz, T. M. (2009). Impact of diagnosticity on the adequacy of models for cognitive diagnosis under a linear attribute structure: A simulation study. *Journal of Educational Measurement, 46,* 450-469.

de la Torre, J., & Lee, Y. S. (2010). A note on the invariance of the DINA model parameters. *Journal of Educational Measurement, 47,* 115-127.

de la Torre, J., & Minchen, N. (2014). Cognitively diagnostic assessments and the cognitive diagnosis model framework. *Psicologia Educative, 20,* 89-97.

de la Torre, J., Hong, Y., & Deng, W. (2010). Factors affecting the item parameter estimation and classification accuracy of the DINA model. *Journal of Educational Measurement, 47,* 227-249.

DiBello, L. V., & Stout, W. (2007). Guest editors' introduction and overview: IRT-based cognitive diagnostic models and related methods. *Journal of Educational Measurement, 44,* 285-291.

Embretson, S. E. (1994). Applications of cognitive design systems to test development. In C. R. Reynolds (Ed.), *Cognitive assessment: A multidisciplinary perspective* (pp. 107-135). New York: Plenum Press.

Embretson, S. E. (1998). A cognitive design system approach to generating valid tests: Application to abstract reasoning. *Psychological Methods, 3,* 380-396.

Ericsson, K. A. (2006). Protocol analysis and expert thought: Concurrent verbalizations of thinking during experts' performance on representative tasks. In K. A. Ericsson, N. Charness, R. R. Hoffman, & P. J. Feltovich (Eds.). *The Cambridge handbook of expertise and expert performance* (p.p. 223-241). New York, NY: Cambridge University Press.

Haser, Ç., & Ubuz, B. (2002). Kesirlerde kavramsal ve işlemsel performans [Conceptual and Procedural Performance in Fractions]. *Eğitim ve Bilim, 27(126)*, 53-61.

Hiebert, J. & Lefevre, P. (1986). Conceptual and procedural knowledge in mathematics: An introductory analysis. In J. Hiebert (Ed), *Conceptual and procedural knowledge: The case of mathematics* (pp. 1-27). Hillsade, NJ: Lawrence Erlbaum Associates.

Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement, 25,* 258-272.

Leighton, J. P. (2004). Avoiding misconception, misuse, and missed opportunities: The collection of verbal reports in educational achievement testing. *Educational Measurement: Issues and Practice, 23*(4), 6-15.

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,* 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*, 205-237.

Mack, N. K. (1995). Confounding whole-number and fraction concepts when building on informal knowledge. *Journal for Research in Mathematics Education, 26,* 422- 441.

Maris, E. (1995). Psychometric latent response models. *Psychometrika, 60*, 523-547.

Maris, E. (1999). Estimating multiple classification latent class models. *Psychometrika, 64,* 187-212.

Misquitta, R. (2011). A review of the literature: Fraction instruction for struggling learners in mathematics. *Learning Disabilities Research and Practice, 26,* 109-119.

Ni, Y. J. (1999). The understanding of the meaning and nature of fraction of Grade fifth and sixth. *Psychological Development and Education, 11*, 26-30.

No Child Left Behind (NCLB) Act of 2001, 20 U.S.C. § 6301 et seq. (2002).

Pellegrino, J. W., Baxter, G. P., & Glaser, R. (1999). Addressing the "two discipline" problem: Linking theories of cognition and learning with assessment and instructional practices. In A.Iran-Nejad & P. D.Pearson (Eds.), *Review of Research in Education* (pp. 307-353). Washington, DC: American Educational Research Association.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.Rproject.org/.

Revelle, W. (2015). "Psych: Procedures for psychological, psychometric, and personality research". In R package Version 1.5.6. Retrieved from http://personalityproject.org/r/psych.manual.pdf.

Rupp, A. A., & Templin, J. L. (2008). Unique characteristics of diagnostic classification models: A comprehensive review of the current state-of-the-art. *Measurement, 6,* 219-262.

Son, J. W. (2012). A cross-national comparison of reform curricula in Korea and the US in terms of cognitive complexity: The case of fraction addition and subtraction. ZDM *The International Journal on Mathematics Education, 44,* 161-174.

Son, J., & Crespo, S. (2009). Prospective teachers' reasoning about students' nontraditional strategies when dividing fractions. *Journal of Mathematics Teacher Education, 12,* 236-261.

Tatsuoka, K. K. (1984). *Analysis of errors in fraction addition and subtraction problems*. Report No. NIE-G-81-0002). Urbana, IL: University of Illinois, CERL.

Tatsuoka, K. K. (1990). Toward an integration of item-response theory and cognitive error diagnosis. In N. Frederiksen, R. Glaser, A. Lesgold, & M. Shafto (Eds.), *Diagnostic monitoring of skill and knowledge acquisition* (pp.453-488). Hillsdale, NJ:Erlbaum.

Tirosh, D. (2000). Enhancing prospective teachers' knowledge of children's conceptions: The case of division of fractions. *Journal for Research in Mathematics Education, 31,* 5-25.

Van de Walle, J., Karp, K. S., & Bay-Williams, J. M. (2013). *Elementary and middle school mathematics methods: Teaching developmentally.* New York, NY: Allyn and Bacon.

Vinner, S., Hershkowitz, R., & Bruckheimer, M. (1981). Some cognitive factors as causes of mistakes in the addition of fractions. *Journal for Research in Mathematics Education, 12,* 70-76.