Evaluating Content Validity of an Instrument to Assess Teachers' Practice of Differentiated Instruction (IATPDI)

Tshering Dorji and Pelden Nima

DOI: https://doi.org/10.17102/bjrd.rub.10.2.006

Abstract

The present study developed and evaluated the content validity of a survey questionnaire titled, "Instrument to Assess Teachers' Practice of Differentiated Instruction (IATPDI)." The items of IATPDI questionnaire were adapted from the pre-existing survey questionnaires and existing literature on differentiated instruction. To quantify the content validity of IATPDI questionnaire, a pre-developed survey questionnaire with 40 items was emailed to two established content experts to: (1) rate each item in terms of its clarity and relevancy to the measured domain; (2) evaluate which item should be deleted or revised; and (3) get recommendation about whether additional items are needed to adequately tap the domain of interest. The Delphi method was used to collect the data. Subsequently, two rounds of Delphi process culminated in a revised and refined IATPDI questionnaire with 40 simplified items divided into four sections. The content validity of IATPDI questionnaire was quantified by calculating the content validity index (I-CVI and S-CVI) and modified kappa statistics, which indicated the high content validity of the 40 items. Taken together, the results show that the IATPDI questionnaire is a content valid instrument.

Keywords*: differentiated instruction, content validity, content validity index*

Introduction

The traditional one-size-fits-all teaching approach no longer meets the needs of today's diverse learners (Burkett, 2013). Indeed, students in today's general education classroom are diverse and take on many forms (Costley, 2012) such as students with different readiness levels, interests, or learning profiles (Tomlinson, 1999, 2001). Succinctly, readiness "refers to the student's prior knowledge, understanding, and skill related to a particular sequence of learning" (Corley, 2005, p. 13) which vary from lesson to lesson or activity to activity. Interest refers to "a child's affinity, curiosity, or passion for a particular topic or skill" (Tomlinson, 1999, p. 11). Learning profile refers to the student's preferred mode of learning (Thiessen, 2012) which is influenced by learning style, grouping preference, and environmental preference (Hall, 2009). Therefore, failing to cater to the diverse needs of learners could certainly hinder and impede their

learning. For instance, previous research suggested that students get frustrated when instruction and learning activities are higher than their readiness level and bored, disappointed, and lose interest when instruction and learning activities are below their readiness level (Dreeszen, 2009; Tomlinson et al., 2003; Valiande & Koutselini, 2009). Consequently, in either situation, students are not able to learn effectively, and unless help is provided, they lose motivation to learn (Dreeszen, 2009). So, it is evident that teachers have to differentiate their instruction in order to meet students' diverse learning needs.

Worldwide, differentiated instruction has proven to be a successful method of teaching diverse students, ever since an influential concept of it was developed by Tomlinson in 1999. Differentiated instruction is a way of teaching and learning (Fisher, 2015; Roberts & Inman, 2013) in which teachers provide multiple avenues to what students learn (content), how they learn it (process), and how they portray what they have learned (product) embracing students as individuals with different readiness level, interests, and learning profiles (Tomlinson, 2001). Different avenues to learning are provided by planning instruction strategically (Corley, 2005), and employing a myriad of research-based didactic and pragmatic strategies and activities that have proven to be successful in addressing diverse students' needs. Further, its principles are rooted in many renowned educational theories on how diverse students learn best (Alberta Education, 2010; Corley, 2005; Hall, 2002; Hobson, 2008; Koeze, 2007; Sherman, 2008) such as Lev Vygotsky's sociocultural theory with key emphasis on Zone of Proximal Development (ZPD), Piaget's constructivist theory, Howard Gardner's theory of multiple intelligences, brain-based learning theory, universal design for learning, and Bloom's taxonomy. Together, these theories undergird effective planning and execution of differentiated instruction (Burkett, 2013; Subban, 2006), which would culminate in scaffolding students' learning.

Interestingly, several studies have reported that effective implementation of differentiated instruction in the mixed ability classrooms increases students' engagement in learning (Beecher & Sweeny, 2008; Hall, 2009; Koehler, 2010; Santangelo & Tomlinson, 2009; Tieso, 2001; Tomlinson & McTighe, 2006; White, 2015), promotes motivation to learn (Fenner & Sydor, 2010; Martin & Pickett, 2013; Massaad & Chaker, 2020; Santangelo & Tomlinson, 2009), exerts a positive attitude towards learning (Beecher & Sweeny, 2008; Chamberlin & Powers, 2010; Karadag &Yasar, 2010), and even minimize behavioural problems or referrals (Cusumano & Mueller, 2007; Greene, 2011; Lewis & Batts, 2005; Waterhouse, 1990, as cited in Visser, 1998) regardless of their differences in learning characteristics.

Subsequently, association between differentiated instruction and students' academic achievement were examined in science (Abigail & Ebele, 2013; Ferrier, 2007; Graham, 2009; Pablico et al., 2017; Sondergeld & Schultz, 2008; White, 2015), reading comprehension and language (Aliakbari & Haghigh, 2014; Beecher & Sweeny, 2008; Boges, 2015; Cusumano & Muelier, 20007; Fisher et al., 2002; Servilio, 2009), and

Mathematics (Amadio, 2014; Beecher & Sweeny, 2008; Butler & Lowe, 2010; Cannon, 2017; Ogunkunle & Henrietta, 2014; Magayona & Tan, 2016; Tieso, 2001, 2002, 2005). These studies found that students exposed to differentiated instruction performed significantly higher than those taught with a time-honored one-size-fits-all traditional method.

However, teachers rarely or occasionally differentiate instruction in their classrooms (Moon et al., 2002; Smith & Humpert, 2012) despite the huge benefits conferred by differentiated instruction in enhancing learning. To this end, Lavania and Nor(2020) found that the most common challenge teachers face in its implementation is a lack of knowledge of differentiated instruction. The authors reviewed nineteen studies published from 2014 to 2019 on the challenges teachers face in implementing differentiated instruction. Hence, it is imperative to provide extensive professional development or training to teachers to help them gain, expand and refine their knowledge of differentiated instruction. This process will enable them to respond appropriately to the individual needs of the students based on their readiness, interests, and learning profiles. To develop effective professional development or training, it is first necessary to assess: (1) teachers' current level of practice of differentiated instruction; (2) teachers' familiarity with instructional and management strategies used to differentiate instruction; (3) factors that help or hinder the implementation of differentiated instruction; and (4) resources and professional development teachers` need to enhance their knowledge and understanding about differentiated instruction. Although there are few survey questionnaires to decipher the aforementioned issues (Adlam, 2007; Crowder, 2011; James, 2009; Logan, 2011; McLean, 2010; Siam & Al-Natour, 2016; Whipple, 2012), to our knowledge, none of these studies have examined and reported on its content validity.

According to Rubio et al. (2003), content validity should be the first psychometric test that should be conducted whenever a new research instrument is developed. Similarly, Sireci (1998) acknowledged the importance of evaluating the content validity of a survey questionnaire and asserted that it is imperative in portraying the evidence of other forms of validity, particularly, construct validity. Validation of content validity checks whether items of data collection instrument are relevant to and representative of the targeted construct of interest (Davis, 1992; Nunnally & Bernsterin, 1994). Furthermore, Rubio et al. (2003) categorized content validity as face validity or logical validity. According to them, face validity involves evaluating whether or not each item measures what it is supposed to measure, "on its face" and logical validity involves a more rigorous process, such as using a panel of experts to evaluate if each item measures the targeted construct it is designed to measure.

To date, quantification of content validity of the aforementioned existing survey questionnaires has never been performed through content validity index (CVI) and modified Kappa statistics, which are two quantitative approaches to quantify content validity of research instrument. Therefore, to address this limitation, the present study

developed an assessment tool titled “Instrument to Assess Teachers’ Practice of Differentiated Instruction (IATPDI) and evaluated its content validity. To collect data for quantification of content validity a modified Delphi method was employed.

Methodology

To estimate the content validity of the IATPDI, this study followed a two-stage process (Development and Judgment/Quantification Stage) advocated by Lynn (1986). The first stage, or “Development Stage,” involved the development of the research instrument. The second stage, or Judgment/Quantification Stage,” involved the content validation of the IATPDI questionnaire. The content validation process was carried out using the modified Delphi method, originally conceived by Dalkey and Helmer in the early 1950s at the Rand Corporation (Goodman, 1987). Briefly, the Delphi method is a consensus-building methodology based on the idea that the collective opinion of identified experts on a specific topic can yield better results than the limited view of an individual (Nworie, 2011). It is used whenever policies, plans or ideas have to be based on informed opinions and judgment of experts and practitioners (Yousuf, 2007). Moreover, it confers an advantage when factors like time and cost make it unlikely or impossible to convene experts in one physical location (Yousuf, 2007) because Delphi’s strength lies in its ability to be administered without face-to-face confrontation among the experts (Grant & Kinney, 1992; Hsu & Stanford, 2017). Since the Delphi method is based on written information and does not require the physical presence of the experts, the method facilitates international, email- or internet-based execution of studies. Likewise, its efficacy for establishing the content validity of the survey questionnaire in research was already demonstrated by earlier studies (Grant & Kinney, 1992; Parratt et al., 2015; Perroca, 2011; Van der Schaaf & Stokking, 2011), which makes it one of the effective tools in content validating research instrument.

In this study, we modified the traditional Delphi method by sending a pre-developed survey questionnaire to the established content experts through email to rate or evaluate each item in terms of its clarity and relevancy to the measured domains. In the traditional Delphi method, the first round is used to generate a list of ideas, or issues toward which consensus is desired and begins with an open-ended questionnaire. The questionnaire rounds are stopped when an acceptable level of consensus is reached and it depends on the level of consensus desired by the researcher (Grant & Kinney, 1997). In the present study, due to time constraints of the experts, only two rounds of the Delphi method were performed to complete the content validation process of the IATPDI questionnaire as opposed to three or more rounds in the traditional Delphi method to reach the consensus among the experts (Grant & Kinney, 1997; Green, 2014; Hasson, Keeney, & McKenna, 2000; Worthen & Sanders, 1987, as cited in Yousuf, 2007).

Development Stage: Development of IATPDI Questionnaire

All the items of the IATPDI questionnaire were adapted from a variety of sources: preexisting survey questionnaires used in earlier studies (Adlam, 2007; Crowder, 2011; James, 2009; Logan, 2011; McLean, 2010; Siam & Al-Natour, 2016; Whipple, 2012), conceptual or theoretical model (e.g., Lev Vygotsky's socio-cultural theory with key emphasis on Zone of Proximal Development (ZPD), Piaget's constructivist theory, Howard Gardner's theory of multiple intelligences, brain-based learning theory, universal design for learning, and Bloom's taxonomy of thinking and learning), and existing literature on differentiated instruction (Corley, 2005; Fisher, 2015; Hall, 2002, 2009; Roberts & Inman, 2013; Tomlinson, 1999, 2001; Tomlinson et al., 2003). Databases including Google Scholar, ERIC, PubMed, Scopus, Web of Science, ScienceDirect, Directory of Open Access Journals (DOAJ), JSTOR, SpringerLink, Alberta Journal of Educational Research, and ResearchGate were searched for existing literature on "differentiated instruction," and "survey questionnaire on differentiated instruction". Additionally, articles and books published by Association for Supervision and Curriculum Development (ASCD), "a key player in advocating a shift to differentiation" (Subban, 2006, p. 935) were also searched. ASCD promotes strategies and tools to help teachers around the world to differentiate instruction in response to diverse learner needs.

Our pre-developed IATPDI questionnaire consists of four sections: A, B, C, and D with a total of 40 items (available at https://shorturl.at/hsCFJ). Section A consists of 7 questions concerning demographic information about the teachers, including gender, level of education, and years of teaching experience, subject areas and grades taught. Data from this part of the questionnaire can be used to investigate whether there is a relationship between demographic characteristics of teachers with the level of implementation of differentiated instruction.

Section B, Implementation of Differentiated Instruction, consists of 35 items categorized into 4 domains of differentiated instruction namely assessment, content, process, and product. Items were designed to be rated on a 4-point Likert scale: (1) Never do this, (2) Seldom (infrequently/rarely), (3) Sometimes (on certain occasions/in certain circumstances) (4) Often (frequently/many times). Data from this part of the questionnaire can be used to assess teachers' level of differentiation of content, process, and product to meet diverse students leaning needs.

Section C and D of the instrument were adapted from Adlam (2007). The permission to adapt and use her survey questionnaire was obtained. Under Section C, item 36 can be used to ask about whether teachers are familiar with various differentiated instructional and management strategies with a two-point dichotomous choices (Yes/No). Likewise, item 37 can be used to ask about their frequency of use of these strategies in their classrooms with a four-point response options of Never

(coded as 0), Once a week (coded as 1), Twice a week (coded as 2), and Thrice or more times a week (coded as 3).

Under section D, item 38, 39 and 40 can be used to ask about factors that help or hinder the implementation of differentiated instruction in the classrooms. The participants can be asked to choose an appropriate response from the menu of options and then check all that apply to them. Finally, item number 40 can be used to ask about the resources that teachers would be willing to use in enhancing their knowledge and understanding about differentiated instruction and choose an appropriate response from the menu of options and then check all that apply to them.

Judgment/Quantification Stage: Evaluating the Content Validity of IATPDI Questionnaire

The content validity of the IATPDI questionnaire was assessed by following the process described by several researchers (Lynn, 1986; Polit & Beck, 2006; Polit et al., 2007; Rubio et al., 2003) through two rounds of the modified Delphi method.

Participants for Delphi Method

We recruited two established content experts having research interests in the field of differentiated instruction. 'Established experts' is defined as a person who exhibits experience in differentiated instruction and/or differentiated instruction research as evidenced by the number of publications of relevant books or papers in that field. We invited six experts through email following the recommendation of Lynn (1986). In the email, we described the purpose and importance of the study and explained the time required to complete the Delphi rounds. The experts were given one week to respond to our request to participate in the study. Of the six invited experts, only two consented to participate in the study (Table 1).

Table 1: Details of the Content Experts Selected for Assessing Content Validity

Designation of Experts	Organization	Qualification
Dr Carol Le Lant	Flinders University	PhD
Dr. Rebecca Saunders	Murdoch University	PhD

Administration of Delphi Method

Two rounds of Delphi method were executed to complete the content validation of IATPDI. The first round involved e-mailing consent letter, information cover letter, content evaluation form and pre-developed IATPDI questionnaire to two content

experts. The information cover letter explained the purpose of the study, the reasons for selecting the content expert, a brief description of IATPDI questionnaire and an explanation of the content evaluation procedure. This round focused on finding out which item should be deleted or revised, or get advice about whether additional items are needed to adequately tap the domain of interest, or to find out if aspects of the construct are represented by the items in correct proportions (Polit et al., 2007). Then, each expert was asked to rate individual item of the pre-developed IATPDI questionnaire in terms of its clarity and relevancy to the construct being measured using a 4-point ordinal scale adapted from Davis (1992) as shown in Table 2. Specifically, if the items are rated 1, 2, and 3, experts were asked to state which item should be deleted or revised, or recommend additional items that can adequately tap the domain of interest (Polit et al., 2007) in the space provided below each domain in the content evaluation form. A week's time was given to complete and return the evaluation form through email.

Table 2: Criteria for Measuring Face and Content Validity

Relevancy	**Clarity**
1= the item is not relevant to the measured domain	1= Not clear
2= the item is somewhat relevant to the measured domain	2= Items need some revision
3= the item is quite relevant to the measured domain	3= Clear but need minor revision
4= the item is very relevant to the measured domain	4= Very clear

Subsequently, item rated 1, 2, and 3 on clarity index, due to usage of obscure vocabulary, ambiguous sentence, and jargons, were modified according to their feedback. Thereafter, based on their ratings on relevancy of each item to the measured domains, content validity index for each item and for whole scale was calculated as discussed under judgment/quantification stage. Using the calculated CVI, modified kappa statistic was also calculated. I-CVI values and kappa coefficients were used to decide on item revisions or deletions in the IATPDI questionnaire. Items not meeting acceptable level of I-CVI and kappa coefficient were modified according to the feedbacks of the experts. Finally, the revised IATPDI questionnaire along with calculated CVI (I-CVI and S-CVI) and modified kappa coefficients were returned to the two experts for their final judgment; due to the experts' time constraints, only two rounds of Delphi method were executed.

Content Validity Index (CVI): Quantification of Content Validity

The content validity of the instrument was quantified by calculating content validity index (CVI) for all individual items (I-CVI) (Lynn, 1986) and the overall scale (S-CVI)

(Polit & Beck, 2006). CVI is an index that expresses the degree of agreement (Polit & Beck, 2006). I-CVI quantifies the extent of agreement among experts on each item (Ozer et al., 2013) and measures the content validity of individual item while S-CVI calculates the content validity of the overall scale (Polit & Beck, 2006; Rodrigues, 2017; Shrotryia & Dhanda, 2019). To calculate I-CVI, the proportion of experts who rated either 3 or 4 was divided by the total number of experts (Polit & Beck, 2006; Polit et al., 2007; Rubio et al., 2003) as shown in Table 3. Subsequently, items with an I-CVI lower than 0.80 were revised and those with very low values were eliminated. Specifically, in case of two experts, an I-CVI of 0.80 is required to establish the content validity of research instrument (Davis, 1992).

Thereafter, two different indices of S-CVI: (1) S-CVI/Universal Agreement (S-CVI/UA) and (2) S-CVI/Average (S-CVI/Ave) were calculated to ensure the content validity of the overall scale. To calculate S-CVI/UA, all S-CVI equal to 1.00 was added and divided by the total number of items. Similarly, to calculate, S-CVI/ Ave, the sum of all I-CVI was divided by the total number of items (Polit & Beck, 2006; Polit et. al., 2007).

Modified Kappa Statistic: To Remove Random Chance Agreement

One criticism of CVI, or proportion agreement, is that the I-CVI values may be inflated (values higher than it should be or than is reasonable) because of possibility of chance agreement (Banerjee et al., 1999; Polit & Beck, 2006; Wynd et al., 2003). To overcome this criticism, modified Kappa statistics (K^*) was calculated since it adjusts each I-CVI values for chance agreement (Polit et al., 2007). It was calculated as

$$K^* = (\text{I-CVI} - P_c)/(1-P_c),$$

Where $P = [N!/A!(N - A)!] \times 0.5_c^N$ is the probability of chance agreement; where N = number of experts in the panel; A = number of experts in the panel who agree that the item is relevant; and $N! = N \times (N - 1) \times \ldots \times 3 \times 2 \times 1$(Fleiss et al., 2003; Larsson et al., 2015: Polit et al., 2007; Shrotryia & Dhanda, 2019).

Since the value of kappa can range between 0 and 1, the standards described in Fleiss (1981) and Cicchetti and Sparrow (1981) were applied to interpret whether the obtained K^* is fair (.40-.59), good (.60-.74), or excellent (>.74). Items with kappa value lower than .74 were revised.

Findings

Table 3 shows the relevance ratings on 40 items by two Delphi experts, calculated I-CVI indices, and probability of chance agreement, kappa coefficients and their interpretations.

Table 3: Rating on 40-Items of Seven Domains by 2 Experts: Items Rated 3 or 4 on 4-Point Relevance Scale

Item No.	Expert 1 (A)	Expert 2 (B)	Experts in Agreement (A+B)	I-CVI Indices	$P_c = \left[\frac{N!}{A!\,(N-A)!}\right] \times 0.5^N$	Kappa coefficients $K^* = \frac{\text{I-CVI} - P_c}{1 - P_c}$	Interpretation of K*
1.	✓	✓	2	1	0.25	1	Excellent
2.	-	-	0	0	0.25	-0.33	Negative
3.	✓	✓	2	1	0.25	1	Excellent
4.	✓	✓	2	1	0.25	1	Excellent
5.	✓	✓	2	1	0.25	1	Excellent
6.	✓	✓	2	1	0.25	1	Excellent
7.	✓	✓	2	1	0.25	1	Excellent
8.	✓	✓	2	1	0.25	1	Excellent
9.	✓	✓	2	1	0.25	1	Excellent
10.	✓	✓	2	1	0.25	1	Excellent
11.	✓	✓	2	1	0.25	1	Excellent
12.	✓	✓	2	1	0.25	1	Excellent
13.	✓	✓	2	1	0.25	1	Excellent
14.	✓	✓	2	1	0.25	1	Excellent
15.	✓	✓	2	1	0.25	1	Excellent
16.	✓	✓	2	1	0.25	1	Excellent
17.	✓	✓	2	1	0.25	1	Excellent
18.	✓	✓	2	1	0.25	1	Excellent
19.	✓	✓	2	1	0.25	1	Excellent
20.	✓	✓	2	1	0.25	1	Excellent
21.	✓	✓	2	1	0.25	1	Excellent
22.	✓	✓	2	1	0.25	1	Excellent
23.	✓	✓	2	1	0.25	1	Excellent
24.	✓	✓	2	1	0.25	1	Excellent
25.	✓	✓	2	1	0.25	1	Excellent
26.	✓	✓	2	1	0.25	1	Excellent
27.	✓	✓	2	1	0.25	1	Excellent
28.	✓	✓	2	1	0.25	1	Excellent
29.	✓	✓	2	1	0.25	1	Excellent
30.	✓	✓	2	1	0.25	1	Excellent
31.	✓	✓	2	1	0.25	1	Excellent
32.	✓	✓	2	1	0.25	1	Excellent
33.	✓	✓	2	1	0.25	1	Excellent

34.	✓	✓	2	1	0.25	1	Excellent
35.	✓	✓	2	1	0.25	1	Excellent
36.	✓	✓	2	1	0.25	1	Excellent
37.	✓	✓	2	1	0.25	1	Excellent
38.	✓	✓	2	1	0.25	1	Excellent
39.	✓	✓	2	1	0.25	1	Excellent
40.	✓	✓	2	1	0.25	1	Excellent
Average I-CVI = 0.98 S-CVI/UA = 0.98 S-CVI/Ave = 0.98							

N = number of experts, A = number of experts in the panel who agree that the item is relevant, and $N! = N \times (N - 1) \times ... \times 3 \times 2 \times 1$

Content Validity Index for Individual Items

Table 3 shows the content validity index for forty items. Thirty-nine items (97.5%) were rated as 3 (the item is quite relevant to the measured domain) and 4 (the item is highly relevant to the measured domain) on the relevancy index by both the experts. The obtained I-CVI value for these items was equal to 1. On the other hand, one item, that is, item number 2 under the Assessment domain was rated as 2 (the item is somewhat relevant to the measured domain) on the relevancy index by one of the experts. Therefore, the obtained I-CVI value for this item was equal to 0.00.

Content Validity Index of the Overall Scale

As seen in Table 3, the content validity indices of the overall scale (both S-CVI/ UA and S-CVI/Ave) are equal to 0.98, which indicates moderate content validity of the IATPDI questionnaire.

Kappa Statistic

Table 3 shows the value of kappa coefficients (column 7) for individual items. The obtained kappa coefficients of thirty-nine items (97.5%) were 1, while kappa coefficient of one item, that is, item number 2 was - 0.33.

Discussion

This study developed and established the content validity of IATPDI questionnaire, designed to (1) assess teachers' level of implementation of differentiated instruction; (2) examine teachers' familiarity with instructional and management strategies used to differentiate instruction as well as how often they use these strategies in their classrooms; (3) identify the factors that facilitate or hinder the implementation of differentiated instruction in their classrooms, and (4) collect information on resources and professional development that teachers` would be willing to use and attend to enhance their knowledge and skills about differentiated instruction.

We followed a two-stage process (Development and Judgment/Quantification Stage) advocated by Lynn (1986). The first stage, or "Development Stage," involved the development of IATPDI questionnaire. All the initial 40 items for this instrument were adopted form the pre-existing survey questionnaires and literatures on differentiated instruction. The second stage, or Judgment/Quantification Stage," involved the assessment of content validity of IATPDI questionnaire. Content validity was evaluated to provide evidence about whether or not each item of IATPDI questionnaire measures the targeted construct it is designed to measure (Davis, 1992; Nunnally & Bernsterin, 1994; Rubio et al., 2003). According to Rubio et al. (2003), the assessment of content validity is the first psychometric test that should be completed whenever a data collection instrument is developed to collect, measure, and analyze data related to the construct of interest.

For a judgment/quantification stage, we used Delphi method. Two rounds of Delphi method were executed. In the first round, a pre-developed survey questionnaire was sent to two established content experts to rate and evaluate each item in terms of its clarity and relevancy to the measured constructs. To quantify their rating, two experts were asked to use 4-point ordinal scale adapted from Davis (1992) as shown in Table 2. Thereafter, content validity of IATPDI questionnaire was quantified by calculating CVI (I-CVI and S-CVI) based on the rating of two experts as detailed in Table 3.

As seen in the Table 3, the I-CVI values of thirty-nine-items (97.5%) were 1. According to Lynn (1986), when there are five or fewer experts, items I-CVI must be 1.00. Therefore, the obtained I-CVI value of 1.00 for each item in the present study is well supported as there were only two content experts to validate the instrument. Davis (1992) further supports the notion that for new instruments, investigators should seek 80% or better consensus among experts.In contrast, I-CVI value of only one item, that is, item number 2 under Assessment domain was 0.00 as both experts rated 2 on relevancy and clarity index on 4-point ordinal scale. They rated item number 2 as "the item is somewhat relevant to the measured domain" on relevancy and "item needs some revision" on clarity. Subsequently, changes were made to improve its clarity and relevancy to the measured construct. For example, the item "To assess student's readiness level, I administer pre-tests, question students about their background

knowledge, or use KWL charts (charts that ask students to identify what they already know, what they want to know, and what they have learned about a topic)" was modified as "To assess each student's readiness level, I pre-test them, question them about their background knowledge, use KWL charts (*charts that ask students to identify what they already know, what they want to know, and what they have learned about a topic*), concept inventories (*multiple choice or short answer tests*), concept map activities, etc."

Further, both the S-CVI/UA and S-CVI/Ave were 0.98 (98%) which indicated high content validity of the overall scale. Scale developers often use a criterion of 0.80 as the lower limit of acceptability for an S-CVI (Polit et al., 2007).

Since CVI utilizes proportion agreement, it has been criticized by many researchers (Polit et al., 2007; Wynd et al., 2003). For instance, Polit et al. (2007) asserted that because of its failure to adjust for chance agreement there is risk of getting inflated values (values higher than it should be or than is reasonable). Therefore, modified Kappa statistic was computed using the calculated I-CVI values (detailed in Table 3) for each item to address the issues of chance agreement (Polit et al., 2007). In addition, Wynd et al. (2003) argued that, though both proportion agreement (CVI) and Kappa coefficient of the agreement provide quantifiable methods for evaluating the judgement of content experts, Kappa offers additional information beyond proportion agreement because it removes random chance agreement. As expected, item receiving lower kappa coefficients were consistent with items having lower I-CVI ratings (Wynd et al., 2003). Parallel to I-CVI values, kappa coefficients for thirty-nine items (97.5%) were 1, except for item number 2 which was -0.33. Considering the standards described in Fleiss (1981) and Cicchetti and Sparrow (1981), a value of Kappa equal to 1 implies perfect agreement between two experts, indicating an excellent content validity of IATPDI of the present study. According to Wynd et al. (2003), a value of kappa coefficient equal to negative implies disagreement between two experts. Instead of deleting item number 2, we modified to improve its clarity and relevancy to the measured construct based on two experts' suggestions as discussed above.

Overall, two Delphi rounds were executed and in these two rounds, minor changes to the items of IATPDI questionnaire was done based on the comments and suggestions received from the two Delphi experts except for item number 2. Two rounds of Delphi process resulted in a refined IATPDI questionnaire with 40 simplified and refined items divided into 4 sections (available at https://shorturl.at/hsCFJ). All the 40 items of the pre-developed IATPDI questionnaire were retained. The quantification of content validity through CVI (I-CVI and S-CVI) and Kappa statistics based on the rating of two experts as detailed in Table 3 indicated high content validity of each item. These results suggest that IATPDI questionnaire is a content valid instrument.

Conclusion

In summary, following a two-stage process (Development and Judgment/ Quantification Stage) advocated by Lynn (1986), this paper developed a survey questionnaire titled "Instrument to Assess Teachers' Practice of Differentiated Instruction" (IATPDI), and assessed its content validity. We established the IATPDI questionnaire's content validity using a modified Delphi method. The calculated content validity index (I-CVI and S-CVI) and coefficients of modified Kappa statistics based on the rating of two experts evidenced high content validity of each item, and therefore, can be a useful tool for conducting further research.

Nevertheless, this study could recruit only two experts for Delphi method to complete content validation process of IATPDI questionnaire. The quantification of content validity through content validity index (I-CVI and S-CVI) and modified Kappa statistic were according to the ratings of two experts. This may have resulted in a high content validity index (I-CVI and S-CVI) and kappa coefficients. Therefore, future studies with more number of experts, preferably more than two, are recommended to address the issue. Another possible limitation of this study is the lack of evaluation of its reliability (test-retest and internal consistency reliability) and other validity (construct and criterion-related validity).

References

Abigail, M. O., & Ebele, C. O. (2013). Effect of differentiated instruction on the academic achievement of Nigerian secondary school biology students.*International Research Journals, 4*(7), 555-560.

Adlam, E. (2007). *Differentiated instruction in the elementary school: Investigating the knowledge elementary teachers possess when implementing differentiated instruction in their classrooms* [Unpublished Master's thesis]. University of Windsor.

Education, A. (2010). Making a difference: Meeting diverse learning needs with differentiated instruction. Edmonton: Alberta Education. Retrieved from http://education.alberta.ca/media/1234045/makingadifference_2010.pdf

Aliakbari, M., & Haghighi, J. K. (2014). Impact of differentiated instruction strategies and traditional-based instruction on the reading comprehension of Iranian EFL students.*Research in Applied Linguistics, 5*(1), 109-129.

Amadio, R. (2014). *Differentiated instruction in secondary mathematics* [Unpublished master's thesis]. University of Wisconsin.

Banerjee, M., Capozzoli, M., Mcsweeney, L., & Sinha, D. (1999). Beyond kappa: A review of interrater agreement measures. *The Canadian Journal of Statistics, 27*(1), 3-23.

Beecher, M., & Sweeny, S. M. (2008). Closing the achievement gap with curriculum enrichment and differentiation: One school's story.*Journal of Advanced Academic, 19*(3), 502-530. Retrieved from https://files.eric.ed.gov/fulltext/EJ810785.pdf

Boges, C. E. (2015). *The effects of differentiated instruction on the achievement score of struggling fourth grade readers* [Unpublished doctoral dissertation].Walden University.

Burkett, J. A. (2013). *Teacher perception on differentiated instruction and its influence on instructional practice* [Unpublished doctoral dissertation]. Oklahoma State University.

Butler, M., & Lowe, K. V. (2010). Using differentiated instruction in teacher education.*International Journal for Mathematics Teaching & Learning,* 1-10. Retrieved from http://www.cimt.org.uk/journal/butler.pdf

Cannon, M. A. (2017). *Differentiated mathematics instruction: An action research study* [Unpublished doctoral dissertation]. University of South Carolina.

Chamberlin, M., & Powers, R. (2010). The promise of differentiated instruction for enhancing the mathematical understandings of college students. *Teaching Mathematics and Its Applications: An International Journal of the IMA*, *29*(3), 113-139.

Corley, M. A. (2005). Differentiated instruction: Adjusting to the needs of all learners. *Focus on Basics,7*(C), 13-16.

Cicchetti, D. V., & Sparrow, S. S. (1981). Developing criteria for establishing interrupter reliability of specific items: Applications to assessment of adequate behaviour. *American Journal of Mental Deficiency. 86,* 127-137.

Costley, K. C. (2012). Who are today's students in a diverse society? Arkansas Tech University.Retrieved from https://files.eric.ed.gov/fulltext/ED528453.pdf

Crowder, I. G. (2011). *Teachers' perception and practices of differentiated instruction at an innovative middle school for gifted and talented students* [Unpublished doctoral dissertation]. University of Georgia.

Cusumano, C., & Mueller, J. (2007). How differentiated instruction helps struggling students. *Leadership, 36*(4), 8-10.

Danzi, J., Reul, K., & Smith, R. (2008). *Improving student motivation in mixed ability classrooms using differentiated instruction* [Unpublished master's thesis]. Saint Xavier University & Pearson Achievement Solutions, Inc.

Davis, L. L. (1992). Instrument review.Getting the most from a panel of experts. *Applied Nursing Research, 5*(4), 194-197.

Dreeszen, J. L. (2009). *The impact of differentiation on the critical thinking of gifted readers and the evolving perspective of the fifth grade classroom teacher* [Unpublished doctoral dissertation]. Kansas State University.

Drost, E. A. (2011). Validity and Reliability in social science research. *Educational Research and Perspectives, 38*(1).105-123.

Fenner, D., Mansour, S., & Sydor, N. (2010).*The effects of differentiation and motivation on students' performance* [Unpublished master's thesis]. Saint Xavier University.

Ferrier, A. M. (2007). *The effects of differentiated instruction on academic achievement in a second-grade science classroom* [Unpublished doctoral dissertation]. Walden University.

Fisher, R. (2015). Competencies needed by teachers of gifted and talented students. (Position statement). South Kitsap School District. Retrieved from http://shorturl.at/bdtwI

Fisher, D., Frey, N., & Williams, D. (2002). Seven literacy strategies that work. *Educational Leadership, 60*(3), 70-73.

Fleiss, J. (1981). Statistical methods for rates and proportions (2nd ed.). New York: John Wiley.

Goodman, C. M. (1987). The Delphi technique: a critique. *Journal of Advanced Nursing, 12*, 729-734. Retrieved from https://onlinelibrary.wiley.com/doi/epdf/10.1111/j.1365-2648.1987.tb01376.x

Graham, K. J. (2009). *Mandated differentiated instruction effectiveness examined* [Unpublished doctoral dissertation]. Walden University.

Grant, J. S., & Kinney, M. R. (1992). Using the Delphi technique to examine the content validity of nursing diagnoses. *Nursing Diagnoses, 3*(1), 12-22.

Green, R. A. (2014). The Delphi technique in educational research.*SAGE Open,* 1-8. Retrieved from https://journals.sagepub.com/doi/pdf/10.1177/2158244014529773

Greene, T. M. (2011). *Differentiated instruction: Reducing behavioural issues in the art room* [Unpublished master's thesis]. LaGrange College.

Hall, B. (2009). Differentiated instruction: Reaching all students. *Research into Practice Mathematic. R*etrievd from https://assets.savvas.com/asset_mgr/current/201033/MatMonHall.pdf

Hall, T. (2002). Differentiated instruction. Wakefield, MA: National Center on Accessing the General Curriculum. Retrieved from https://www.dr-hatfield.com/educ342/Differentiated_Instruction.pdf

Hasson, F., Keeney, S., & McKenna, H. (2000). Research guidelines for the Delphi survey technique. *Journal of Advanced Nursing, 32*(4), 1008-1015.

Hobson, M. L. (2008). *An analysis of differentiation strategies used by middle school teachers in heterogeneously grouped classrooms.* [Unpublished master's thesis]. University of North Carolina.

Hsu, C., & Sandford, B. A. (2007). The Delphi technique: Making sense of consensus. *Practical Assessment, Research, and Evaluatio*n, *12*(10), 1-8.

James, D. (2009). Differentiated instruction: One school's survey analysis. *The Corinthian*, *10*(13), 169-190. Retrieved from https://kb.gcsu.edu/cgi/viewcontent.cgi?article=1075&context=thecorinthian

Kalpan, R. M., Bush, J. W., & Berry, C. C. (1976). Health status: Types of validity and the index of well-being. *Health Services Research, 1*(4), 478-507.

Karadag, R., & Yasar, S. (2010). Effects of differentiated instruction on students' attitudes towards Turkish courses: An action research. *Procedia Social and*

Behaviorla Sciences, 9, 1394-1399. Retrieved from https://files.eric.ed.gov/fulltext/ED554550.pdf

Koehler, S. (2010). *Effects of differentiating for readiness, interest and learning profile on engagement and understanding* [Unpublished master's thesis].St. John Fisher College.

Koeze, P. P. (2007). *Differentiated instruction: The effect on student achievement in an elementary school* [Unpublished doctoral dissertation]. Eastern Michigan University.

Lavania, M., & Nor, F. B. M. (2020). Barriers in differentiated instruction: A systematic review of the literature. *Journal of Critical Reviews, 7*(6), 293-297.

Lewis, S., & Batts, K. (2005). How to implement differentiated instruction? Adjust, Adjust, Adjust: North Carolina project begins with encouragement from administrators. *Journal of Staff Development, 26*(4), 26-31.

Logan, B. (2011). Examining differentiated instruction: Teachers respond. *Research in Higher Education Journal*, *13*, 1-14.

Lynn, M. R. (1986). Determination and quantification of content validity.*Nursing Research, 35*, 382-385.Retrieved from https://doi.org/10.1097/00006199-198611000-00017

Magayon, V. C., & Tan, E. B. (2016). Learning mathematics and differentiated instruction in the Philippines: A Phenomenographical study on struggles and successes of grade 7 students. *International Journal of Educational Studies in Mathematics, 3*(3), 1-14. Retrieved from https://dergipark.org.tr/en/download/article-file/397462

Martin, M. R., & Pickett, M. T. (2013). *The effect of differentiated instruction on motivation and engagement in fifth-grade gifted math and music students* [Unpublished master's thesis]. Saint Xavier University.

Massaad, M., & Chaker, L. Y. A. (2020). The influence of differentiated instruction on Lebanese students' motivation, knowledge, and engagement.*International Journal of English Literature and Social Sciences, 5*(3), 755-774. doi: 10.22161/ijels.53.31

McLean, V. M. (2010). *Teacher attitude toward differentiated instruction in third grade language arts* [Unpublished doctoral dissertation].University of Southern Mississippi.

McCrae, R. R., Kurtz, J. E., Yamagat, S., & Terracciano, A. (2011). Internal consistency, retest reliability, and their implications for personality scale validity. *Personality and Social Psychological Review, 15*(1), 28-50. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2927808/

Moon, T. R., Callahan, C.M., Tomlinson, C. A., Miller, E. M. (2002). Middle School Classrooms: Teachers` Reported Practices and Student Perceptions. *National Research Centre on the Gifted and Talented.*

Nworie, J. (2011). Using the Delphi technique in educational technology research.*TechTrends, 55*(5). 24-30. Retrieved from https://snazlan.files.wordpress.com/2015/11/using-the-delphi-technique-in-educational-technology-research-by-nworie-2.pdf

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric Theory* (3rd Ed.). Mcgraw-Hill: New York.

Ogunkunle, R. A., & Henrietta, O. A. (2014). Effect of differentiated instructional strategies on students' retention in geometry in FCT senior secondary schools, Abuja, Nigeria.*Global Journal of Educational Research, 13*, 1-7. Retrieved from https://www.ajol.info/index.php/gjedr/article/view/169846

Ozer, I., Fitzgerald, S. M., Sulbaran, E., Garvey, D. (2014). Reliability and content validity of an English as a Foreign Language (EFL) grade-level test for Turkish primary grade students. *Procedia – Social and Behavioral Sciences, 112*, 924-929.

Parraca, J. A., Olivares, P. R., Carbonell-Baeza, A., Aparicio, V. A., Adsuar, J. C., &Gusi, N. (2011). Test-retest reliability of biodex balance SD on physically active old people. *Journal of Human Sport and Exercise, 6*(2), 444-451.

Parratt, J. A., Fahy, K. M., Hutchinson, M., Lohmann, G., Hastie, C. R., Chaseling, M., & O'Biren, K. (2016). Expert validation of a teamwork assessment rubrics: A modified Delphi study. *Nurse Education Today, 36*, 77-85.

Pablico, J., Diack, M., & Lawson, A. (2017). Differentiated instruction in the high school science classroom: Qualitative and Quantitative analysis. *International Journal of Learning, Teaching and Educational Research, 16*(7), 30-54.

Polit, F. D., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations.*Research in Nursing & Health, 29,* 489-497. Retrieved from https://onlinelibrary.wiley.com/doi/epdf/10.1002/nur.20147

Polit, F. D., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations.*Research in Nursing, 30*, 459-467. Retrieved from https://onlinelibrary.wiley.com/doi/epdf/10.1002/nur.20199

Roberts, J. L., & Inman, T. F. (2013). *Teacher's survival guide: Differentiating instruction in the elementary classroom*. Prufrock Press.

Rodrigues, I. B., Adachi, J. D., Beattie, K. A., & MacDermid, J. C. (2017). Development and validation of a new tool to measure the facilitators, barriers and preferences to exercise in people with osteoporosis.*BMC Musculoskeletal Disorders, 18*(1), 1-9. Doi: 10.1186/s12891-017-1914-5

Rodriguez, A. (2012). *An analysis of elementary school teachers' knowledge and use of differentiated instruction* [Unpublished doctoral dissertation]. Olivet Nazarene University.

Rubio, D. M., Berg-Weger, M., Tebb, S. S., Lee, E. S., & Rauch, S. (2003). Objectifying content validity: Conducting a content validity study in social work research. *Social work Research,* 27(2), 94-104.

Santangelo, T., & Tomlinson, C. A. (2009). The application of differentiated instruction in postsecondary environments: Benefits, challenges, and future directions. *International Journal of Teaching and Learning in Higher Education, 20*(3), 307-323.

Servilio, K. L. (2009). You get to choose! Motivating students to read through differentiated instruction.*Teaching Exceptional Children Plus, 5*(5), 2-11.

Sherman, W. (2008). Differentiated instruction: A review of literature. Richmond Metropolitan Educational Research Consortium (MERC)

Shrotryia, V. K., & Dhanda, U. (2019). Content validity of assessment instrument for employee engagement.*SAGE Open,* 1-7. Retrieved from https://doi.org/10.1177/2158244018821751

Siam, K., & Al-Natour, M. (2016). Teacher's differentiated instruction practices and implementation challenges for learning disabilities in Jordan. *International Education Studies, 9*(12), 167-181. Retrieved from https://files.eric.ed.gov/fulltext/EJ1121601.pdf

Sireci, S. G. (1998). The construct of content validity.*Social Indicators Research, 45*, 83 – 117.

Smith, R., & Humpert, W. (2012). Differentiated instruction in small schools. *Teaching and teacher education, 28*(8), 1152-1162.

Sondergeld, T. A., & Schultz, R. A. (2008). Science, standards, and differentiation: It really can be fun! *Gifted Child Today*, *31*(1), 34-40.

Subban, P. (2006). Differentiated instruction: A research basis. *International Education Journal,7*(7), 935-947. Retrieved from https://files.eric.ed.gov/fulltext/EJ854351.pdf

Tavakol, M., Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education, 2,* 53-55.

Thiessen, A. (2012). *Differentiated Physical Learning Environment* (Unpublished master's thesis). Dordt College.

Tieso, C. (2001). Curriculum: Broad brushstrokes or paint-by-the numbers? *The Teacher Educator, 36*(3), 199-213. Retrieved from https://www.tandfonline.com/doi/abs/10.1080/08878730109555264

Tieso, C. (2002). The effects of grouping practices and curriculum practices on intermediate students' mathematics achievement scores.*The National Research Center on the Gifted and Talented.*https://nrcgt.uconn.edu/wp-content/uploads/sites/953/2015/04/rm02154.pdf

Tieso, C. (2005). The effects of grouping practices and curricular adjustments on achievement.*Journal for the Education of the Gifted, 29*(1), 60-89.

Tomlinson, C. A. (1999). *The* differentiated classroom: Responding to the needs of all learners. Alexandria, VA: Association for Supervision and Curriculum Development.

Tomlinson, C. A. (2001). *How to differentiate instruction in mixed-ability classrooms*, (2nd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.

Tomlinson, C. A., Brighton, C., Hertberg, H., Callahan, C. M., Moon, T. R., Brimijoin, K., & Reynolds, T. (2003). Differentiating instruction in response to student readiness, interest, and learning profile in academically diverse classrooms: A view of literature. *Journal for the Education of the Gifted, 27*(2/3), 119-145.

Tomlinson, C. A. (2005). Travelling the road to differentiation in staff development.*Journal of Staff Development,26*(4), 8-12.

Tomlinson, C. A., & McTighe, J. (2006). Integrating differentiated instruction & understanding by design: Connecting content and kids. Alexandria, VA: Association for the Supervision of Curriculum Development.

Valiandes, S., & Koutselini, M. (2009). *Application and evaluation of differentiation instruction in mixed ability classrooms.* 4th Hellenic Observatory PhD Symposium, LSE. [Symposium].University of Cyprus.

Van der Schaaf, M. F., & Stokking, Karel, M. (2011). Construct validation of content standards for teaching. *Scandinavian Journal of Educational Research, 55*(3), 273-289.

Vaske, J. J., Beaman, J., Sponarski, C. C. (2016). Rethinking internal consistency in Cronbach's alpha. *Leisure Sciences,*1-11.

Visser, J. (1998). *An examination of differentiation* [Unpublished official degree thesis]. University of Birmingham.

Waltz, C., & Bausell, R. B. (1983). *Nursing research: Design, statistics, and computer analysis.* Philadelphia: F. A. Davis Company

Whipple, K. A. (2012). *Differentiated instruction: A survey study of teacher understanding and implementation in a southeast Massachusetts school district* [Unpublished doctoral dissertation]. Northeastern University, Boston, Massachusetts.

White, D. M. (2015). *Differentiated instruction in the science classroom: Student perception, engagement, and learning* [Unpublished master's thesis]. Montana State University.

Wynd, C. A., Schmidt, B., & Schaefer, M.A. (2003). Two quantitative approaches for evaluating content validity. *Western Journal of nursing research, 25*(5), 508-518

Yousuf, M. I. (2007). Using experts' opinion through Delphi technique.*Practical Assessment, Research & Evaluation*, *12*(4),1-8.

Yusoff, M. S. B. (2019). ABC of content validation and content validity index calculation.*Education in Medicine Journal, 11*(2), 49-54. Retrieved from https://doi.org/10.21315/eimj2019.11.2.6

About the Authors

Tshering Dorji is currently working as a teacher at Zhemgang Central School, Zhemgang Dzongkhag. He obtained his bachelor's degree in Physical Science (Physics and Chemistry) from Sherubtse College, Kanglung and Postgradutate Diploma in Education from Samtse College of Education (Physics), Samtse. Additionally, he pursued MEXT Teacher Training Program from Kagawa University, Japan. He is a MEXT fellow. His research interests include curriculum and instructional studies, assessment and intervention in science education, and technology in education.

Pelden Nima is currently working as a teacher at Wangbama Central School, Thimphu Dzongkhag. He obtained his bachelor's degree in Biotechnology from Bangalore University, India and Postgradutate Diploma in Education from Samtse College of Education (Biology), Samtse. He pursued his master's degree in Agricultural Sciences from Nagoya University, Japan. He is a MEXT fellow. His research interests include how nutritional factors affect pubertal onset in mammals, ovulation and reproduction in mammals. Besides, he is also keenly interested in exploring how learning triggers neural changes in the brain.