

“It Puts us in our Students’ Shoes”: Listening to Voices from Teacher Candidates on Their Test-Taking Experience

Liying Cheng* and Jia Ma

Assessment and Evaluation Group, Faculty of Education, Queen’s University, Kingston Ontario Canada

Abstract: This study explores teacher candidates’ experiential learning through their test-taking experiences while attending a Bachelor of Education (B. Ed.) program. Eighty-four written reflections by teacher candidates taking a mid-term course examination on classroom assessment practices were analyzed. Major themes emerging from these reflections on the test-taking experience are related to validity concepts of construct representation, construct-irrelevant variance, relevance, and fairness. The study reveals that the test-taking experience could be valuable to teacher candidates in their learning of classroom assessment practices and in their understanding of the issues in test taking that may influence test performance. This, in turn, could potentially provide teacher candidates with a direct framework for their future classroom assessment practices, by which they may support their own future students.

Keywords: Test-taking experience, classroom assessment practices, classroom assessment tools, teacher education.

1. INTRODUCTION

Increasing educational expectations and accountability measures throughout North American jurisdictions have resulted in a growing use of large-scale assessments in schools (Earl, 1995; Mazzeo, 2001). Perhaps the most extreme example of this can be found in the US 2002 ‘No Child Left Behind’ Act, but recent examples in Canada include the Grade 3 Alberta Provincial Achievement test, the Ontario Provincial Assessment of Reading, Writing and Mathematics - Primary Division (Grades 1–3) and Junior Division (Grades 4–6), the Ontario Secondary Schools Literacy Test (OSSLT). In Alberta, grade 3 students who fail the provincial achievement test must write a supplemental examination in Grade 4. In Ontario, successful completion of the OSSLT, or the Ontario Secondary School Literacy Course (OSSLC) implemented for students who have failed the OSSLT, is a requirement of graduation. With ‘No Child Left Behind’, schools and districts have until 2014 to ensure that all students, with few, if any, student exemptions, meet educational expectations. Schools must document their annual progress and those with less than satisfactory performance may face a variety of sanctions. In Ontario, with the publishing of the OSSLT results, schools, teachers, and students are under a great deal of pressure as the impact posed to themselves into the day-to-day classroom (Babcock, 2006; Cheng, Klinger, & Zheng, 2007; Doe, Cheng, Fox, Klinger, & Zheng, 2011; Forrest, 2004; Zheng, Klinger, Cheng, Fox, & Doe, 2011).

Within this educational climate, pre-service teacher education programs need to seriously address how we better prepare future teachers to face the increasing accountability challenges in order to avoid them to accept or reject such large-scale testing and assessment at the face level and to support them to conduct quality classroom assessment in their future classrooms. Teacher candidates need to understand the critical issues in relation to this accountability framework, to gain assessment knowledge in both large-scale testing context and classroom assessment setting, and more importantly, to acquire critical skills of using assessment tools and implementing assessment practices during their teacher education program and their teaching practicum (Hill, Cowie, Gilmore, & Smith, 2010). However, many studies have demonstrated that classroom teachers lack assessment literacy in their teaching practice (DeLuca & Klinger, 2010; Mertler, 2005; Zhang & Burry-Stock, 2003). These studies call for the necessity for teacher candidates to understand the fundamental concepts of and have the knowledge and skills to conduct assessment *for* learning and assessment *of* learning¹. Although there exist suggestions on what teacher candidates should know about assessment and how they use assessment, assessment literacy training for teacher candidates varies greatly on type (compulsory or optional), content (assessment as the sole focus or assessment embedded in other courses), and length (short course

*Address correspondence to this author at the Faculty of Education, Queen’s University, 511 Union Street, Kingston, Ontario, K7M 5R7, Canada; Tel: +001 613 5336000 ext. 77431; E-mail: liying.cheng@queensu.ca

¹Assessment *of* learning refers to those assessments that happen after learning has occurred to determine whether learning has happened. They are used to make statements of student learning status at a point in time. Assessment *for* learning refers to the process of seeking and interpreting evidence for use by students and their teachers to decide where the students are in their learning, where they need to go, and how best to get there.

or course followed by long practicum) (see Hill *et al.*, 2010, for a review).

In short, for most pre-service teacher education programs, how to equip teacher candidates with assessment knowledge to respond to increasing educational accountability context is under exploration (DeLuca, Klinger, Searle, & Shulha, 2010; Hill *et al.*, 2010; Popham, 2011).

2. EXPERIENTIAL LEARNING

Studies have investigated how teacher candidates learn to use assessment knowledge and skills through teaching practicum (DeLuca *et al.*, 2010; Geen, Bassett, & Douglas, 2001; Graham, 2005) - a typical means by which such learning occurs. This can be referred to as experiential learning (see Hildreth, 1951; Kolb, 1984; Korthagen, 2001), because experience (i.e., engaging in tasks and activities in program and in practicum) plays a central role in the holistic integrative learning process, where it also combines perceptions (i.e., continuous reflections on their learning along with their practicum experience) with teacher cognition and behaviours. Experiential learning provides the richness of learning (Lave & Wenger, 1991), valuable reflections on that learning (Schön, 1983), and better supports teacher candidates to bridge theory, experience, and their teaching practicum.

Researchers in cognitive psychology have been paying attention to how individuals make their choices. Some cognitive psychologists argue that individuals' past experiences will affect their future decision-making processes to a great extent. For example, Li, Mayhew, and Kourtzi (2009) showed that learning from past experiences changed the circuitry in individuals' brains, thus individuals can categorize what they have observed in order to make a decision or carry out appropriate actions. Juliusson, Karlsson, and Gärling (2005) also indicated past experience influenced the decisions individuals make in the future. This study indicates that when something positive results from a decision, individuals are more prone to decide in a similar way, given a similar situation. On the other, individuals tend to avoid repeating past mistakes (Sagi & Friedland, 2007). However, Juliusson *et al.* (2005) pointed out that it did not necessarily mean that future decisions made based on past experiences were necessarily the best decisions. Following the perspectives from these psychologists, it can be argued that, teacher candidates' previous assessment (e.g., test-taking) experiences during their teacher

education program are likely to influence their current and future assessment practices. Therefore, it is important for teacher education programs to support teacher candidates' learning through experiences and reflections on their previous experiences.

In addition to experiential learning of assessment in teacher candidates' practicum, assessment courses in teacher education program can also offer opportunities for experiential learning, for example, having teacher candidates taking an actual test. The reason of implementing such experiential learning is that examinations play an important role in assessment practices, in both large-scale setting and classroom setting. Currently, tests are still a very common assessment method in schools (DeLuca & Klinger, 2010) and in universities (Cheng, Rogers, & Wang, 2008). Teacher candidates need to learn the knowledge (e.g., underlying concepts like construct, reliability, and validity) and skills (e.g., test design and construction) if they want to use tests/examinations in their future classroom assessment practices.

Within the current accountability context in North America and globally, all large-scale assessments require students to write tests to judge whether they have mastered the content knowledge corresponding to their grade level. However, when students take large-scale assessments, they are always confronted with issues that may influence their test performance but not necessarily related to their content knowledge (such as anxiety), which are identified as construct-irrelevant variance in educational measurement research (Haladyna & Downing, 2004). One way of reducing construct-irrelevant variance in order to increase valid interpretation of students' test scores is that teachers integrate strategies (such as how to lower test anxiety) in their classroom instructions. Therefore, teacher candidates need to understand what issues may potentially influence students' test performance and test-taking and how these issues interfere with students' test-taking.

3. TEACHER CANDIDATES AS TEST-TAKERS

Educational assessment literature has increasingly pointed out the importance of collecting data from test-takers to support validation evidence of the test (Cheng & DeLuce, 2011). Given the extent of the consequences and influences from both large-scale testing and classroom assessment on instruction, there is a critical need to ensure that testing practices yield valid data about student achievement and performance

(Kane, 2002; Moss, 2007; Stobart, 2003). Moss, Girard, and Haniford (2006) argue that validation studies must include multiple stakeholder perspectives (e.g., test takers, test users) in order to expose sources of evidence that would otherwise stand to invalidate test inferences. Previous research has demonstrated strong evidence of test validation from the perspective of test developers (Bachman, 2007). However, validity evidence from the perspectives of test-takers has been limited (Cohen, 2006; Hamp-Lyons, 2000). Validation evidence from test-takers should include an analysis of “how test-takers interpret test constructs and the interaction between these interpretations, test design, and accounts of classroom practice” (Fox & Cheng, 2007, p. 9).

Despite the fact that testing including test taking is a very important component of teacher candidates' knowledge sets, only a few studies investigated teacher candidates' attitudes towards testing and testing-related practices (Beghetto, 2005; Green, 1992; Kher-Durlabhji, Lacina-Gifford, Carter, & Jones, 1995). For example, Green (1992) found out pre-service teachers, compared with in-service teachers, had more favorable attitudes towards large-scale testing than classroom testing; Beghetto (2005) indicated that pre-service teachers with positive past testing experience tended to use testing more often than those with negative experience; and Kher-Durlabhji and colleagues (1995) revealed that pre-service teachers were less able to judge the appropriateness of score-increasing strategies in the grey area on ethical-to-unethical continuum. Although these studies contributed to the understanding of pre-service teachers' perceptions on testing, the experience and perceptions of teachers being test-takers is still under study. Although Beghetto examined pre-service teachers' past testing experience, the undefined “past” and “testing” lead into some questions to make interpretation of the results limited. For example, how long ago should the experience counted as “past”? Does the “past” include all the period from primary to secondary education? What kind of tests, classroom tests or large-scale tests, should be included in “testing experience”? All of these questions make it worthwhile to continue the investigation of teacher candidates as test takers in more specific context, such as test-taking experience in their teacher education program.

Because of the central role that experience plays in the learning process of teacher candidates and because of the important accountability role that tests play in schooling, it is worthwhile to employ test-taking

experience as a way in assessment education for teacher candidates. In addition, examining test-taking experience can add to our understanding of teacher candidates' perceptions on testing, especially from the perspective of them being test-takers. Therefore, the purpose of this study is to explore the test-taking experience of teacher candidates taking a mid-term course examination and to conduct test evaluation of the written examination. This study aims to better understand whether test taking could be a valuable approach and experience for teacher candidates in learning assessment knowledge and skills from their own perspective. Specifically, the following research questions were examined:

1. How do teacher candidates perceive their test-taking experience? Is test-taking a valuable experience for teacher candidates?
2. How do teacher candidates perceive their evaluation of the written examination? Is test evaluation a valuable experience for teacher candidates?

4. METHODOLOGY

4.1. Participants

The participants were 84 teacher candidates in a one-year Bachelor of Education (B. Ed.) pre-service teacher education program at a Canadian university. These participants were from two cohorts of teacher candidates who took the same course on classroom assessment practices in two recent consecutive academic years. This course typically examined the roles and impacts of assessment on teaching and learning, with a focus on the philosophical foundations and the technical aspects of assessment. Teacher candidates were required to develop expertise in the planning and construction of assessment tasks while integrating learner-centered and growth-oriented approaches to assessment practices. All teacher candidates in this course over the two cohorts were invited to participate in the study. These participants were pursuing their B. Ed. degree in addition to another bachelor degree in their subjects of teaching. They were roughly between 22 and 35, and with a rough equal number of male and female teacher candidates. They were predominantly born and raised around the province of Ontario Canada, and all intended to become educators and/or classroom teachers at the secondary level; with teachables ranging from English to technology education. All of them had completed

their first two of four teaching practicums prior to taking this course, thus they had some experience of classroom assessment practices.

4.2. Instrument and Data Collection

Data were collected from the participants' written reflections on their experience of taking a mid-term course examination. As a way of experiential learning and with a purpose to better prepare the teacher candidates with assessment literacy, this examination was designed and administered to explore the role of test-taking and the role of test evaluation as an assessment learning experience for teacher candidates. The mid-term examination was one of the three required assessment components of this course – the other two were written assignments for the teacher candidates to demonstrate their knowledge and competence in key aspects of classroom assessment practices. This written examination consisted of 25 items of a variety of item formats, e.g., multiple choice and short answer questions, which addressed both the philosophical and technical aspects of assessment and evaluation, e.g., validity, reliability, fairness, scoring, and reporting, etc. The examination was administrated as an open-book examination. The performance of this examination and their participation in test-taking reflection (plus two other assignments together) contributed to the participants' completion of the course, which was evaluated with a Pass and Fail scoring system and laid out explicitly in the course outline.

The purposes of this test-taking and test evaluation experience were also to educate the teacher candidates about testing – the quality of the test, test instructions, test items, through taking an actual test. The teacher candidates were informed that the purposes of this test-taking experience were:

- To review the concepts learned so far in the course;
- To reflect on the test-taking experience;
- To subsequently evaluate the test.

All the participants spent 35 minutes on the test, and then marked each other's examination paper in pairs, afterwards they reflected on test-taking experience and conducted the test evaluation. They evaluated the test's layout, directions, formats, and items. They employed, but were not limited to, a list of the dimensions of test evaluation based on the criteria

discussed in the course textbook (Taylor & Nolan, 2008). In the end, they wrote written reflections on their test-taking and their test evaluation.

4.3. Data Analysis

Although existing studies have investigated teacher candidates' perceptions of testing (e.g., Green, 1992; Kher-Durlabhji, *et al.*, 1995), few have investigated how their test-taking experience may influence their own learning as teacher candidates (e.g., Beghetto, 2005). Grounded theory was thus adopted to analyze data as a way of exploring the data (Corbin & Strauss, 2008). The grounded theory can keep the researchers focus on data analysis, thus making the analysis more original and useful (Charmaz, 2006). More importantly, it enables the researchers to study the *processes* (which is the focus of this study—the teacher candidates' experiential learning of taking a test) and explain what is happening in the processes and in the *field setting* (where the teacher candidates are experiencing test taking) (see Charmaz, 2006). An inductive and comparative approach was employed in the data analysis to generate codes. Specifically, the following steps were used: reviewing all 84 written reflections collected from two cohorts; identifying descriptive segments with participants' perceptions and opinions; grouping these segments addressing the same issue into one code, labeling the code, counting the number of these segments as frequency of the code, repeating the previous steps, and making a list of all codes. Then, related codes were categorized to form themes (see Table 1).

We have chosen to name the themes based on the wordings used mostly often by the participants to keep them as original as what participants expressed in their written reflections and also the validity framework used in the course textbook (Taylor & Nolan, 2008). The constant comparison analytic method was used to pay attention to codes general across the two cohorts. We found similar results across the two cohorts so the data were reported below accordingly. In addition to manually analyzing the data, the NVivo software was also used to sort out and organize segments under each code, and to tabulate the frequencies of codes and themes as double-checking to ensure the consistency of frequencies counted by the manual analysis. Two researchers coded data to ensure the consistency of coding; and the interrater reliability, which was calculated in this process, was 95%. Where a discrepancy existed in data coding, consensus was

reached between the two researchers through further discussions on code meaning.

Table 1: Themes and Codes Frequencies

Themes and Codes	Frequency
Theme 1—Construct Representation	70
Code 1: Coverage of course content	16
Code 2: Presentation of test content	6
Code 3: Accuracy and clarity	48
Theme 2—Construct-irrelevant Variance	111
Code 4: Test layout	19
Code 5: Test instructions	9
Code 6: Test format	19
Code 7: Test administration	42
Code 8: Test-taking strategies	4
Code 9: Test anxiety	18
Theme 3—Relevance	16
Code 10: Learning opportunity	8
Code 11: Connection and application	8
Theme 4—Fairness	7
Code 12: Perceived fairness	5
Code 13: Availability of resources	2

5. RESULTS

Four themes arose from our analysis: construct representation, construct-irrelevant variance, relevance, and fairness (see Table 1). The participants' perceptions of taking this test will be described briefly first as a prelude to presenting these overarching themes, and then each theme will be elaborated.

Approximately 30% of the participants (25/84) commented on their general perceptions of taking this test. 32% (8/25) had a "neutral feeling" about taking this test, and approximately 40% said the test helped them to review the course contents. As a way of enabling students understand the important ingredients of testing, this course examination was designed as a typical classroom test, and it included various item formats (e.g., multiple-choice, short answer), and was required to be completed on an answer sheet within a designated timeframe. Therefore, this "typical test" made most of the participants "felt like being tested," and they thought that it was quite different from "doing in-class essays". Upon completing this test, 8% of these participants (2/25) said they felt "relief when done" and "better than hoped".

5.1. Construct Representation

Construct representation is a variety of aspects that the participants addressed to judge the quality of this mid-term examination. Specifically, the aspects they addressed included: their general perception of the validity of this examination, whether assessment knowledge and skills are built into this examination, whether the test items are relevant to what participants learned in this assessment course, and whether the technical qualities of this test design are achieved. These aspects match with the content aspect of construct validity that Messick (1996) has synthesized—evidences of content relevance and representativeness as well as technical quality. Therefore, *construct representation* arose as one major theme from the participants' reflection on their test taking and test evaluation experience and this theme will be elaborated in the following aspects: coverage of course contents, presentation of test contents, and accuracy and clarity of test items.

Sixteen participants (19%) (16 [frequency of *Coverage of Course Content*] divided by 84 [the total number of the participants' reflections]) discussed the coverage of course contents. All of these participants commented whether the test contents matched and assessed what they had learned from this course. Some participants found the test covered the materials discussed in class, and was "consistent with teaching instructions, and therefore was "a good way to review what we have learned thus far"; however, some thought the test did not focus on what they learned, and thus the test "did not test our knowledge accurately".

The presentation of the test contents was also addressed (7.1%: 6/84). Some of the participants' evaluation included looking at whether item formats could properly assess their learning. They felt that multiple-choice questions, true and false questions, and short-answer questions, were limited because these items mainly measured their memorization skills rather than what they "really learned". Some participants said that "choosing the best answer type of questions is ineffective, because it does not promote open-mindedness, but merely a memorization or ranking of concepts," and "it would have been nice to demonstrate some more complex skills, i.e., creating a task analysis²."

²Creating a task analysis is one of the required assignment of this course.

More than half of the participants were concerned about the accuracy and clarity of test items and test choices (57%: 48/84). Most said that wording of this test was difficult to understand; and some similar and tricky multiple-choice options made them feel there were more than one correct answer. One participant noted that "The test was frustrating [because] many of the answers for the MC [multiple-choice questions] were hard to distinguish among [each other] and seemed that all said the same thing." For the multiple choice questions that contains "NOT"³ in question stems, some participants said that "there [were] lots of NOT questions" and "[such questions] led to [confusion] and not easy to understand".

Some participants addressed grammatical ambiguity as well as the use of negative and superlative words. One participant reflected that, "the level of grammar used in the test seem[ed] higher than what was covered in the classes, I felt like [I had] to use a dictionary to look up words on the MC [multiple-choice] questions." Some participants also mentioned the negative and superlative words in test items. They felt that "[there were] too many NOTs and BESTs [capitalization in the original reflections] in item statements and options, and [it is better to] have more definite questions and answers", and they felt that answering such test items were time-consuming and difficult.

5.2. Construct-Irrelevant Variance

Construct-irrelevant variance refer to the conditions, variables or factors that were unrelated to the participants' competence or skills assessed in this examination but which may influence their performance on examination (see Haladyna & Downing, 2004; Messick, 1989). Such factors occur with high frequency (see Table 1) in the participants' reflections (111 on construct-irrelevant variance vs. 70 on construct representation). Therefore, *construct-irrelevant variance* was identified as one major theme, including test layout, test instruction, test format, test administration, test-taking strategies, and test anxiety.

Nineteen participants (22.6%: 19/84) expressed the concerns about test layout. They felt that answer sheets and organizations of the test items were hindrance to their test-taking. Some participants commented that there were limited space on the

answer sheets and the structure of the answer sheets was not easy to follow. Some said, "the space given for the written portion was too small, [and I] feel pressured to keep writing to a minimum [length]", and "[I] feel like that I [should] rip the answer sheet off and write on it separately." Organization of test items was also reported as a distractor in participants' test-taking: "it bothers [me] when [I have to] switch between short answer questions and multiple-choice questions." Some participants complained that the difficult items at the beginning made the test seem "very scary".

Some participants (10.7%: 9/84) noted the influence of the test instruction on their test-taking. The extent of clarity and the designated scores in test instructions were reported as their concerns. Participants wanted to know the expected length (e.g., number of words) of answers to short-answer questions and the expected number of choices to multiple-choice questions. One participant said, "The examination was not clear in its expectations. The instructions were to 'circle the answers', [but] it did not say circle only the best answer." Some participants were unclear about whether it was worth of spending time on some questions rather than others because of the relatively small score weights.

More than one-fifth of the participants (22.6%: 19/84) made comments on the test formats. Many of them expressed their own preference over particular test formats. Some liked short-answer questions because these questions allowed them to "reflect and gather information collectively". Others preferred multiple-choice questions and commented positively on the "excellent design" of the written examination because "they activate[d] a memory of what I have learned," and "the MC [multiple choice] format kept my anxiety down because I know the right answer is in front of me." Some participants expected that a variety of test formats should be used to assess their learning: "overall I would have liked to see more examples of different testing formats" and "instead of defining feedback, perhaps [the question should] ask a paragraph describing how to use [feedback]."

Half of the participants (42/84) noted some test administration issues that influenced their test taking. These issues included: (a) open-book test, and (b) test location and food. This examination was open-book and some participants felt it made the questions easily answered by employing their "speedy reference skills" to search in textbooks. Thus they thought this open-book test was not a good assessment strategy

³For example, *What is NOT the correct statement of validity?*

because they were not “encouraged to make efforts to prepare and study before the test”. Some particular events - room change, seating and food during the test-taking were noted as the factors influencing the participants’ test performance. One participant said, “changing room was difficult to adjust as there was a lot going on in the room itself⁴. Seating was not arranged strictly to prevent students from cheating, and some participants found it was a concern because students “can use anyone else’s answers” if they wanted to cheat, especially for the multiple-choice questions. One participant said, “ [It is a problem of] using MC [multiple choice] question[s] and having people sit so close to each other, [so] it is almost unavoidable to see someone else[s] answer sheet [and] it was too easy to cheat.” The participants did enjoy fruit provided during their test-taking: “the oranges help[ed] create an encouraging and more relaxed environment,” and “it is a proven fact that tests should not be taken on an empty stomach.”

Around 5% of the participants (4/84) commented on their test-taking strategies. The strategies included reading thoroughly, searching references, and using clues. They used the words in the questions and choices as clues to find answers, and one said that, “a thorough reading always revealed the correct answer”. They could find clues in the multiple-choice questions, and this open-book test allowed them to employ their reference skills to find answers. Despite the test being open-book, peer scored and on a pass and fail evaluation, some participants (21%: 18/84) still reported test anxiety and pressure. These participants were nervous before the test and felt stressful through the test, even though they were “confident about the materials”. A very small group of participants felt anxious because they felt time for completing the test was insufficient. One participant said, “after receiving the test, I became increasingly stressed to find the answers within the textbook, [I] felt very rushed for the time throughout the test and found myself barely finishing.” Some participants’ anxiety arose from concerns on their test performance. They noted, “[I felt] a bit of pressure to perform given the fact that I had to pass the test to finish the course;” and “the only hesitation was [that] the test was called a ‘mid-term’, [which] usually refer[s] to a large portion of your final mark.” Peer behavior was also mentioned as a reason that increased participants’ anxiety. One participant

said, “[I] felt frustrated near the end of the test (in the last 10 minutes) when people started to finish, went out in the hall, and were talking loudly. [It] bothers me when people do not show the same respect for others still writing the test.” Another felt the same way - “having students leave the classroom increased the anxiety because you did not want to [be] the last one left on your table.”

5.3. Relevance

Relevance is the third theme emerging from the participants’ comments. The participants commented on how this test-taking experience was related to their learning in this assessment course, assessment practices in their future teaching, and their life experience.

About one-fifth of the participants (16/84) reflected on benefits of this test-taking to their assessment learning. Some thought that the test-taking experience helped them understand the course materials, and that the test prompted them to study, making this test more a learning opportunity than an assessment. Participants thought that the benefit was “learn[ing] to assess by writing tests [of this nature].” One participant said, “[I] appreciate being able to use the textbook, [because I] cared about finding the right answer and expressing my thoughts clearly;” and another said, “[it is] a very useful test for understanding the concepts of classroom assessment, [and] I learn some terminology while accomplishing this [test].”

Some participants related their test-taking experience to their future teaching and study. Some commented on the benefits of being test-takers and thought taking this test made them experience what factors might influence their test performance. One participant wrote that, “the test help[ed] us to remember what it was like to take[ing] tests back in high school; *it puts us in our students’ shoes* [emphasis added].” The test-taking experience influenced some participants’ choices of their future classroom assessment practices, e.g., preference of some test formats over others. One participant said, “I have always had trouble with [answering] MC [multiple choice questions]. I will probably use only a few MC questions on tests I [will] write [in the future] because I can’t write them very well, and the questions [that are not written well] would be unfair to my students.” Some participants reflected the benefit of this test taking to their study. One participant found the reason of making mistakes in tests and said, “[by] looking back at the mistakes, I realized I didn’t take the time to review any of my decisions made on

⁴The room change was an unfortunate event due to a mixed up of classrooms between graduate and undergraduate teaching scheduling.

[answering] test [questions].” One participant emphasized the importance of using available resources, and wrote that, “I find having resources and knowing where to find the information is more applicable to [my] study and life experiences.”

5.4. Fairness

Fairness is the fundamental concept in assessment and evaluation. The participants commented on fairness in terms of whether their test performance was an accurate indicator of their learning in this assessment course. Two aspects of test fairness were derived from the participants' written comments: perceived fairness and availability of resources. Generally, most of the participants who commented on the perceived fairness thought the test was fair in terms of test items design, coverage of course materials, and representation of what they had learned in this course. “The test was fair considering what we learned in the lectures, and “[it] seemed fair and tested the material that I thought would be on it.”

Two participants commented on the availability of access to the textbook during their test-taking. Because they had no textbook as reference to answer test questions, they thought that they did not have the equal opportunity to write the test as other students who had textbooks. Although the textbook was a required text for this course, not every teacher candidate purchased the book. “[I] Felt that students without a book were highly disadvantaged because on a few instances there were definitions to be given,” and “[The open-book test] Would only work if all students have a textbook, which is not always possible.” The availability of textbook was reasonably identified as a factor influencing the participants' test performance because this test was administered as an open-book one. The availability of textbook to all the participants could provide equal opportunity to demonstrate their competencies, which was how the participants interpreted fairness in their test-taking.

6. DISCUSSIONS

The written reflections revealed that these teacher candidates evaluated this test by the criteria of assessment *for* and *of* learning as well as major validity framework (Messick, 1989; Moss, Girard, & Haniford, 2006). The test-taking experience provided these teacher candidates with a context and also a process to employ these criteria, and with an opportunity to reinforce these criteria in their assessment mindsets and to help them review important concepts of

classroom assessment. These results highlight the characteristics and value of experiential learning that has been demonstrated by the findings of some previous studies (see Hildreth, 1951; Korthagen, 2001) - the process of learning is entailed more emphasis than the outcomes (Kolb, 1984). In relating to the specific learning process of taking this mid-term examination, the influences, benefits and learning of this test-taking experience are presented to answer the two research questions posed for the study, i.e., is test-taking a valuable experience for teacher candidates from their own perspectives.

First, the reflections on the test taking provided an opportunity for these teacher candidates to experience how test-taking can be used as both assessment *of* learning and assessment *for* learning practices. These teacher candidates clarified some unclear and ambiguous conceptual ideas and reviewed concepts they had learned about classroom assessment. We recognized that teacher candidates presented some contradictory comments about the test, which showed different uptake of teacher candidates' assessment knowledge. It is possible that taking a real-test was a useful way to accentuate some concepts that might lead to such contradictory reflections. One teacher candidate said that some test questions went against course materials presented in class. It was through actually taking the test that this candidate noticed the discrepancies between his/her initial understanding of some concepts presented in class and the potential alternative explanations. The comments on using this test-taking experience to understand and learn concepts of classroom assessment were concrete examples of teacher candidates' assessment learning. In this way, this test-taking experience has provided these teacher candidates with an opportunity to review and reinforce conceptual ideas of classroom assessment.

Second, the test evaluations provided the teacher candidates with an opportunity to examine the technical components of this test and to experience how these technical issues may have influenced their test performance. All the aspects of test evaluation that teacher candidates addressed were based on their perceptions as to whether their learning was assessed accurately by this test. The data from written reflections of this study showed the ambiguity and confusion teacher candidates felt in their test-taking and thus they directed their attention onto some aspects of test design, such as test directions, formatting, and the grammar of test items. When evaluating this test,

teacher candidates realized how technical issues of test design could influence their test-taking and their test performance (e.g., clarity or ambiguity of test items and choices). Because technical quality of assessments is related to construct validity (Messick, 1989, 1996), attention on these technical components and quality that teacher candidates acquired in this test-taking experience may increase their awareness of test quality when they use tests and examinations in their future teaching practice.

Third, this test-taking experience helped these teacher candidates think how to implement different test formats in their future assessment practices. The variety of test formats used in this test provided them with a direct experience of how and what aspects of their learning were assessed. Through taking this test and conducting test evaluation, these teacher candidates had seen the different assessment purposes served by different test formats. The participants reflected in their written reflections that multiple-choice questions were effective to assess understanding of concepts but not effective to promote open-mindedness and to demonstrate more complex skills. Some teacher candidates indicated their preference over certain test formats in their assessment practice based on their test-taking experience. One teacher candidate with negative experience with multiple-choice questions said he would not use multiple-choice questions in his future assessment practice. This is consistent with Beghetto's (2005) findings that teacher candidates with negative past testing experience will use less testing in their teaching practice. Although Bethetto did not specify "past" in his study, it is also reasonable to believe that, the immediate "past" test-taking experience that teacher candidates had in this assessment course may have some influence on their future assessment practices.

Fourth, this test-taking experience provided the teacher candidates with an opportunity to review and understand the potential factors and conditions that could influence test performance (e.g., test layout, test instruction, test administration, and test anxiety). All of their observations and reflections through this test-taking experience could become available resources for their future classroom instructions and assessment. This test-taking experience, as one teacher candidate commented, "helped us to remember what it was like to take[jing] tests back in high school; *it puts us in our students' shoes.*" By experiencing the problems and anxiety encountered through test administration, these teacher candidates had a contextual understanding of

what conditions and variables other than test-takers' academic knowledge/skills could influence actual test performance. With this understanding, these teacher candidates may provide their future students with instructions on how to cope with the factors and conditions that may influence test performance. Such instructions may help teacher candidates better prepare their future students for assessment tasks.

Many studies have shown that teacher candidates could lack the awareness of how assessment would influence their students' learning (Hill, Cowie, Gilmore, & Smith, 2010; Seigel, 2007; Seigel, Wissehr, & Halverson, 2008) because these teacher candidates themselves tend to be successful test-takers when they were students. The attention on teachers' competence in classroom assessment has been rising (Stiggins, 1995) in response to studies demonstrating that classroom teachers lack assessment skills and competence in their teaching practice (Cheng, Rogers, & Wang, 2008; Mertler, 2005; Zhang & Burry-Stock, 2003). One possible strategy to cope with the above phenomena can be, as shown by the findings of this study, to situate teachers and teacher candidates in real tests (even though it is a classroom test) in order to increase their awareness of assessment skills and competence, and to experience what specific assessment skills they need to improve in their classroom assessment practices.

7. CONCLUSIONS

By analyzing the written reflections from 84 teacher candidates taking a course examination, we have empirically examined, from teacher candidates' own perspectives, how they experienced test-taking and test evaluation and whether those experiences benefited these candidates. Within this current educational climate, teacher education programs need to seriously address how to better prepare future teachers in assessment literacy to face the increasing accountability challenges. These teacher candidates need not only to understand the critical issues in relation to this accountability framework but also to gain the knowledge of classroom assessment practices so they can best support their future students. How could these teacher candidates prepare their future students for the increase use of testing without even experiencing taking a test themselves and reflecting on the test and test-taking experience within their teacher education program? Through this test-taking experience, teacher candidates in this study have had the opportunity to reflect on potential issues relating to

test design, test condition, and testing fairness, which shed lights on some aspects that may be absent or abstract from what they have learned in academic courses and in the textbooks. Since teacher education must empower future teachers with critical thinking skills that can enable their continuous growth in their professional lives, the results of this study demonstrate their reflections on this test-taking experience and their test evaluation can contribute to their development of assessment literacy and also potentially contribute to test validation evidence from the perspective of teacher candidates as test-takers. Acknowledging this is a naturalistic and qualitative study conducted within the classroom setting and from the perspectives of the teacher candidates alone, there are a number of limitations of the study. First, this study employed a single data source of written reflections. Additional source of follow-up interviews with these participants would provide more contextual information about their understanding and reflections, which can potentially strengthen the study. Second, we chose to let the teacher candidates to conduct their reflections immediately after taking the test and within the classroom setting. An extended time and away from the classroom testing might have enabled the reflections to be more in-depth and avoid certain sensitive and nuance issues relating to conducting the reflections around their peers and their instructor even though the teacher candidates were not required to identify themselves. Third, we acknowledge the potential limitations of our study that relies exclusively on participants' self reports and one time test-taking experience. Further research should also collect observation data from teacher candidates' practicum or possibly follow them into their future teaching. Future assessment courses could introduce a series of such experiential activities to support teacher candidates' assessment learning. Only then can we fully realize how valuable such an experience has been to these teacher candidates and whether they are able to act on their assessment learning.

Despite its limitations, the study offers an initial and exploratory attempt to add to our understanding of teacher candidates' assessment literacy and how to approach assessment and evaluation in pre-service teacher education program. The knowledge of assessment is perceived to be a wicked problem by many teacher candidates (Shulha *et al.*, 2012), and is of core attention to teacher education community to improve the effectiveness of training teacher candidates on this knowledge set and capacity. This study situated the teacher candidates in an experiential learning opportunity to have them "apply and analyze"

what they "remember and understand" about the conceptual knowledge of assessment. These teacher candidates' reflections presented us potential benefits of using course examination as an option for assessment education for teacher candidates, which is a common assessment method that classroom teachers use with their students but that receives little attention in teacher education, especially assessment education of teachers.

REFERENCES

- Babcock, M. S. (2006). *The Impact of a culture of Large-scale Assessment and High-stakes Results Accountability for elementary principals in one Ontario school board*. Unpublished M. Ed. thesis. Queen's University, Kingston, Ontario, Canada.
- Bachman, L. (2007). What is construct? The dialectic abilities and contexts in defining constructs in language assessment. In J. Fox, M. Wesche, D. Bayliss, L. Cheng, C. Turner & C. Doe (Eds.), *Language Testing Reconsidered* (pp. 41-72). Ottawa, ON: University of Ottawa.
- Beghetto, R. A. (2005). Preservice teachers' self-judgments of testing taking. *Journal of Educational Research*, 98, 376-380. <http://dx.doi.org/10.3200/JOER.98.6.376-380>
- Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative research analysis*. London: Sage.
- Cheng, L., & DeLuca, C. (2011). Voices from test-takers: Further evidence for test validation and test use. *Educational Assessment*, 16, 104-122.
- Cheng, L., Klinger, D., & Zheng, Y. (2007). The challenges of the Ontario Secondary School Literacy Test for second language students. *Language Testing*, 24(2), 185-208. <http://dx.doi.org/10.1177/0265532207076363>
- Cheng, L., Rogers, T., & Wang, X. (2008). Assessment purposes and procedures in ESL/EFL classrooms. *Assessment & Evaluation in Higher Education*, 33(1), 9 – 32. <http://dx.doi.org/10.1080/02602930601122555>
- Cohen, A. D. (2006). The Coming of age of research on test-taking strategies. *Language Assessment Quarterly*, 3(4), 307-331. <http://dx.doi.org/10.1080/15434300701333129>
- Corbin, J., & Strauss, A. (2008). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory* (3rd ed.). Thousand Oaks, CA: Sage.
- DeLuca, C., & Klinger, D. (2010). Assessment literacy development: identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy and Practice*, 17(4), 419-438. <http://dx.doi.org/10.1080/0969594X.2010.516643>
- DeLuca, C., Klinger, D., Searle, M., & Shulha, L. (2010). Developing a Curriculum for Assessment Education. *Assessment Matters*, 2, 20-42.
- Doe, C., Cheng, L., Fox, J., Klinger, D., & Zheng, Y. (2011). What has experience got to do with it? An exploration of L1 and L2 OSSLT test takers' feelings, perceptions of test performance, and alignment to classroom literacy activities. *Canadian Journal of Education*, 34(3), 68-85.
- Earl, L. (1995). Assessment and Accountability in Education in Ontario. *Canadian Journal of Education*, 20(1), 45-55. <http://dx.doi.org/10.2307/1495051>
- Forrest, M. (2004). *The Impact of Grade 6 Testing on Teachers*. Unpublished M. Ed. thesis. Queen's University, Kingston, Ontario, Canada.
- Fox, J., & Cheng, L. (2007). Did we take the same test? Differing accounts of the Ontario Secondary School Literacy Test by first and second language test-takers. *Assessment in Education: Principles, Policy & Practice*, 14(1), 9-26. <http://dx.doi.org/10.1080/09695940701272773>

- Geen, A., Bassett, P., & Douglas, L. (2001). Preparing student teachers to assess pupil's achievements. *Westminster Studies in Education*, 24(1), 35-43.
<http://dx.doi.org/10.1080/0140672010240104>
- Graham, P. (2005). Classroom-based assessment: Changing knowledge and practice through preservice teacher education. *Teaching and Teacher Education*, 21(6), 607-621.
<http://dx.doi.org/10.1016/j.tate.2005.05.001>
- Green, K. E. (1992). Differing opinions on testing between pre-service and in-service teachers. *The Journal of Educational Research*, 86, 37-42.
<http://dx.doi.org/10.1080/00220671.1992.9941825>
- Haladyna, T. M., & Downing, S. M. (2004). Construct-irrelevant variance in high-stakes testing. *Educational Measurement: Issues and Practices*, 23(1), 17-27.
<http://dx.doi.org/10.1111/j.1745-3992.2004.tb00149.x>
- Hamp-Lyons, L. (2000). Social, professional and individual responsibility in language testing. *System*, 28, 579-591.
[http://dx.doi.org/10.1016/S0346-251X\(00\)00039-7](http://dx.doi.org/10.1016/S0346-251X(00)00039-7)
- Hildreth, G. (1951). The Role of Experiential Learning in the Education of Teachers. *Journal of Teacher Education*, 2(3), 180-184.
<http://dx.doi.org/10.1177/002248715100200305>
- Hill, M., Cowie, B., Gilmore, A., & Smith, L. F. (2010). Preparing Assessment-capable Teachers: What Should Preservice Teachers Know and be Able to Do? *Assessment Matters*, 2, 6-22.
- Juliusson, E. Å., Karlsson, N., & Gärling, T. (2005). Weighing the past and the future in decision making. *European Journal of Cognitive Psychology*, 17(4), 561-575.
<http://dx.doi.org/10.1080/09541440440000159>
- Kane, M. T. (2002). Validating high-stakes testing programs. *Educational Measurement: Issues and Practices*, 21(1), 31-41.
<http://dx.doi.org/10.1111/j.1745-3992.2002.tb00083.x>
- Kher-Durlabhji, N., Lacina-Gifford, L. J., Carter, R. B., & Jones, R. (1995). Pre-service teachers' views on standardized testing practices. *Research in the Schools*, 2(1), 35-40.
- Kolb, D. A. (1984). *Experiential learning: experience as the source of learning and development*. Englewood Cliffs, NJ: Prentice Hall.
- Korthagen, F. A. J. (2001). *Linking practice and theory: The pedagogy of realistic teacher education*. Paper presented at the Annual Meeting of the American Educational Research Association, Seattle: USA.
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. New York: Cambridge University Press.
<http://dx.doi.org/10.1017/CBO9780511815355>
- Li, S., Mayhew, S. D., & Kourtzi, Z. (2009). Learning shapes the representation of behavioral choice in the human brain. *Neuron*, 62(3), 441-452.
<http://dx.doi.org/10.1016/j.neuron.2009.03.016>
- Mazzeo, C. (2001). Frameworks of state: Assessment policy in historical perspective. *Teachers College Record*, 103, 367-97.
<http://dx.doi.org/10.1111/0161-4681.00119>
- Mertler, C. A. (2005). Secondary teachers' assessment literacy: Does classroom experience make a difference? *American Secondary Education*, 33, 76-92.
- Messick, S. (1989). Validity. In R. L. Linn. (3rd Ed.), *Educational measurement* (pp. 13-104). New York: American Council on Education and Macmillan.
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13, 241-256.
<http://dx.doi.org/10.1177/026553229601300302>
- Moss, P. A., Girard, B. J., & Haniford, L. C. (2006). Validity in educational assessment. *Review of Research in Education*, 30, 109-162.
<http://dx.doi.org/10.3102/0091732X030001109>
- Moss, P. A. (2007). Reconstructing validity. *Educational researcher*, 36(8), 470-476.
<http://dx.doi.org/10.3102/0013189X07311608>
- No Child Left Behind Act. (2002). Pub. L. No. 107-10.
- Popham, W. J. (2011). Assessment literacy overlooked: A teacher educator's confession. *The Teacher Educator*, 46(4), 265-273.
<http://dx.doi.org/10.1080/08878730.2011.605048>
- Sagi, A., & Friedland, N. (2007). The cost of richness: The effect of the size and diversity of decision sets on post-decision regret. *Journal of Personality and Social Psychology*, 93, 515-542.
<http://dx.doi.org/10.1037/0022-3514.93.4.515>
- Schön, D. A. (1983). *The Reflective Practitioner: How Professionals Think in Action*. US: The Basic Books, Inc.
- Shulha, L., Klinger, D., Luu, K., Ma, J., & Lam, C. (2012). *Assessment is a Wicked Problem! The Implications for University Instructors and Pre-Service Teachers*. Paper presentation at 2012 Annual Conference of Canadian Society for the Study of Education (CSSE), Waterloo, Canada.
- Seigel, M. A., Wissehr, C., & Halverson, K. (2008). Sounds like success: A framework for equitable assessment. *The Science Teacher*, March, 43-46.
- Siegel, M. A. (2007). Striving for equitable classroom assessments for linguistic minorities: Strategies for and effects of revising life science items. *Journal of Research in Science Teaching* 44(6): 864-881.
<http://dx.doi.org/10.1002/tea.20176>
- Stiggins, R. J. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan*, 77, 238-245.
- Stobart, G. (2003). The impact of assessment: Intended and unintended consequences. *Assessment in Education: Principles, Policy & Practice*, 16, 139-140.
- Taylor, C. S., & Nolen, S. B. (2008). *Classroom assessment: Supporting Teaching and Learning in Real Classrooms* (2nd ed.). New Jersey: Pearson Education, Inc.
- Zhang, Z., & Burry-Stock, J. A. (2003). Classroom assessment practices and teachers' self-perceived assessment skills. *Applied Measurement in Education*, 16, 323-342.
http://dx.doi.org/10.1207/S15324818AME1604_4
- Zheng, Y., Klinger, D., Cheng, L., Fox, J., & Doe, C. (2011). Test-takers' background, literacy activities, and their views of the Ontario Secondary School Literacy Test. *Alberta Journal of Educational Research*, 57(2), 115-136.