Adapting the Critical Thinking Assessment Test for Palestinian Universities

Sami Basha[1,2], Denise Drane[1] & Gregory Light[1]

[1] Searle Center for Advancing Learning and Teaching, Northwestern University, Evanston, USA

[2] Bethlehem Palestine Ahliya University College, Bethlehem, Palestine

Correspondence: Sami Basha, Bethlehem Palestine Ahliya University College, Bethlehem, Palestine. E-mail: basha@forzapace.net

Abstract

Critical thinking is a key learning outcome for Palestinian students. However, there are no validated critical thinking tests in Arabic. Suitability of the US developed Critical Thinking Assessment Test (CAT) for use in Palestine was assessed. The test was piloted with university students in English (n=30) and 4 questions were piloted in Arabic (n=48). Students responded favorably. Scores were comparable with US scores. Only two students found the content problematic. One-hundred-twelve Palestinian faculty reviewed the skills tested by the CAT. There was moderate agreement that they represent critical thinking. Translation of the CAT into Arabic and further study are warranted.

Keywords: assessment, critical thinking, higher education, Palestine

1. Introduction

The globalization of higher education over the last few decades has been accompanied by an enormous range of different kinds of assessment; witness for example the rash of new international university rankings that have appeared since the onset of the present millennium focused on assessing university's relative reputation (Rauhvargers, 2011). At the regional, national and institutional levels there have also been widespread efforts to assess the quality of teaching as part of broader Quality Assurance (QA) initiatives (Bernhard, 2012). Both of these kinds of efforts have expanded quickly to include national systems of Higher Education of all types across the world. In Palestine where the modern university sector only began in the mid 1970's and has been beset by considerable unique obstacles due in large part to an ongoing era of occupation and conflict (Abu-Lughod, 2000; Abu-Saad & Champagne, 2006), a national Accreditation and Quality Assurance Commission (AQAC) was nevertheless established in 2002.

Like other national QA schemes, complex issues lie at the core of the Palestinian QA initiative: questions about what should be measured, how institutions should assess their educational impact, and which kinds of outcomes provide the most reliable information about institutional strength, teaching effectiveness, and quality of student learning. Moreover, a critical tension between assessment for accountability and assessment for improving learning is frequently missing in broader policy debates. Improvements to teaching and learning based on assessments of higher order student learning outcomes and evaluations of programs which use such assessments are even less frequent. Indeed, despite the growth of alternative learning-centered methods of assessing student learning (Light, Cox, & Calkins, 2009), traditional modes of instruction and assessment continue to be the main methods teachers use to assess their students' learning. In a national study of undergraduate teaching practices in Palestine, Cristillo (2009), found that teaching and assessment practices were primarily geared toward lower levels of learning such as rote memorization.

More recently, however, Palestinian commentators have called for universities to focus on the development of rich, lively and interactive environments to encourage higher order student learning outcomes, in particular creative and critical thinking skills (Fasheh, 2014); both for employability (Palestinian National Authority PNA, 2012) but also to meet ongoing challenges of living and learning in the Palestinian context (AbuLaban, 2014). These appeals for the integration of critical thinking into the curriculum have been championed by many Palestinian academics both at the program level where national programs have specified critical thinking skills and aligned learning

activities as key program goals (Basha, 2012) and at the institutional level where the recent development of learning and teaching centers of excellence have identified critical thinking as a major goal of their faculty development initiatives (Daragmeh, Drane, & Light, 2012). Indeed, as a result of a recent, large scale national project focused on developing learning and teaching in Palestinian universities, (Palestinian Faculty Development Project), four such university centers have been established and three national conferences have been organized emphasizing the centrality of learning and the importance of critical thinking in higher education. In addition, in the last year, two national Capacity Building workshops for both faculty and Palestinian trainers of faculty have been held in Jericho in 2014 and Ramallah in 2015 with a special emphasis on critical thinking.

Despite this broad upsurge in interest and activity in critical thinking across disciplines and fields, there has been little to no development on a test for critical thinking that might be similarly used across disciplines. Currently there are no robust (validated and reliable) tests of critical thinking in Arabic which may be used to assess improvement in student critical thinking over time, and which are appropriate to the Palestinian context. Assessment of students' critical thinking using valid and reliable methods is vital to national and institutional efforts to improve critical thinking to a) insure that progress in critical thinking is actually being made, and b) to help identify teaching approaches that lead to the greatest gains in critical thinking. It cannot be assumed that critical thinking skills will improve in students simply because critical thinking has been targeted as a learning outcome. For example, Rawanda has identified critical thinking as key to its national strategy to develop a skilled worked force (MINDEDUC, 2010). However, in a recent study of students at three prestigious universities in Rawanda, Schendel (2015) found that students were not making meaningful gains on a test of critical thinking designed specifically for the Rawandan context.

The purpose of this study was to assess whether one such instrument-the Critical Thinking Assessment Test (Stein & Haynes, 2011)-used extensively in the United States (US) may be appropriate for the Palestinian higher education context. As Schendel (2013) notes, the validity of an assessment in one context does not automatically indicate that the assessment will be valid in another. The aim of the study was to gather data on a) the response of Palestinian students to the test, and b) the response of Palestinian faculty to the critical thinking skills examined on the test.

2. Assessment of Critical Thinking

2.1 The Critical Thinking Assessment Test (CAT)

The Critical Thinking Assessment Test (CAT) is a 15 item, short answer essay test developed in the United States with the support of the National Science Foundation (NSF) to assess critical thinking skills in undergraduate students in the fields of Science, Technology, Math and Engineering (STEM) and related fields. It tests critical thinking across four core domains: a) evaluation of information, b) evaluation of ideas and other points of view, c) learning and problem solving, and d) communication of ideas. It does not test rote memory of information, but rather requires students to exercise higher-order thinking skills such as those on the upper levels of Bloom's Taxonomy of Educational Objectives (1956): application, analysis, synthesis, and evaluation. Specific critical thinking skills assessed by the CAT are listed in Table 1 below. Test questions are based on real-world scenarios. Most require short essay answers which reveal the students' thought processes. The short essay format was chosen because it has been shown to be less racially biased, have higher construct validity and to test more skills in the same question than multiple choice questions (US Department of Education, 2000). Below is a sample disclosed item from the CAT. While the CAT is not a timed test, most students take approximately one hour to complete it.

"A scientist working at a government agency believes that an ingredient commonly used in bread causes criminal behavior. To support his theory the scientist notes the following evidence.

99% of the people who committed crimes consumed bread prior to committing crimes.

Crime rates are extremely low in areas where bread is not consumed.

Do the data presented by the scientist strongly support their theory? Yes No

Are there any other explanations for the data besides the scientist's theory? If so, describe.

What kind of additional information or evidence would support the scientist's theory?

The CAT may be scored by either faculty or graduate students using the detailed scoring rubrics provided with the test. Importantly, it has been shown to be sensitive to course effects (Stein, Haynes, & Redding, 2006). National US norms for performance on the test are available. It has been administered in several hundred colleges and universities across the US and found to be valid and reliable and appropriate for students across all institutional types and levels. In terms of validity, the CAT has been shown to have satisfactory face validity and

criterion validity (Stein Haynes, Redding, Ennis, & Cecil, 2007). Face validity of the CAT was established in a study in which the 12 skills were shown to faculty from a variety of disciplines at the 6 US universities where the CAT was first developed. Agreement was 80 percent or higher across the 12 skills indicating a high degree of face validity. (Stein et al., 2007). CAT scores are moderately correlated with general measures of academic performance such as the SAT (r=0.527), ACT (r=0.599), and Grade Point Average (GPA) (r=0.345). In addition, scores on the CAT are moderately correlated with scores on other measures of critical thinking such as the California Critical Thinking Skills Test (CCTST; r=0.645). Test-retest reliability is acceptable at > 0.80 (Stein et al., 2007). Internal consistency of CAT items is also acceptable (Cronbach's alpha=0.695) suggesting that items are measuring the same general construct (Stein et al., 2007). Finally, the cultural fairness of the CAT has been evaluated in the US and has shown that neither gender, race nor ethnic background are statistically significantly associated with performance (Stein et al., 2007). A cultural Differential Item Functioning (DIF) analysis has also been performed and indicates that there were no items with prevalent cultural bias (Stein & Haynes, 2011).

Table 1. Critical thinking skills assessed by the CAT

Question	Critical Thinking Skill
Q1	Summarize the pattern of results in a graph without making inappropriate inferences.
Q2	Evaluate how strongly correlation-type data supports a hypothesis.
Q3	Provide alternative explanations for a pattern of results that has many possible causes.
Q4	Identify additional information needed to evaluate a hypothesis or a particular explanation of an observation.
Q5	Evaluate whether spurious relationships strongly support a claim.
Q6	Provide alternative explanations for spurious relationships.
Q7	Identify additional information needed to evaluate a hypothesis/interpretation.
Q8	Determine whether an invited inference in an advertisement is supported by information.
Q9	Provide relevant alternative interpretations of information.
Q10	Separate relevant from irrelevant information when solving a real-world problem.
Q11	Analyze and integrate information from separate sources to solve a real-world problem.
Q12	Use basic mathematical skills to help solve a real-world problem.
Q13	Identify suitable solutions for a real-world problem using relevant information.
Q14	Identify and explain the best solution for a real-world problem using relevant information.
Q15	Explain how changes in a real-world problem situation might affect the solution.

This research consists of a series of 3 studies. Studies 1 and 2 focused on students' responses to the CAT and involved students taking the full test in English and a subset of questions in Arabic respectively. The third study was a survey study focused on faculty responses to the skills assessed on the CAT. Methods, procedures and results of each study are reported below.

3. Study 1: Students' Responses to the CAT Test in English

Aims of this first study were to examine English speaking Palestinian students' responses to the English version of the CAT to a) determine if the students could relate to the contexts used in the questions b) assess their comfort with the test and c) determine if any aspects of the test were confusing for them.

3.1 Methods

3.1.1 Participants

The study sample consisted of a convenience sample of 30 students from the faculties of nursing, medicine (n=28) and information technology (n=2) at 2 large, independent, non-governmental Universities in the Palestine. They were invited to participate in the study by 3 of their course professors who were known to the first author (though not at his home institution in Palestine). They were selected to participate in the study because of their excellent English language skills. All had completed their English proficiency course requirements with a grade of at least a B (i.e., a score of 80). All were told that the test was part of a research study and that their test scores would be used only for the research study and would not contribute to their course grade. Characteristics of the students are presented in Table 2. Males and females were equally represented in the sample. They ranged in age from 19-21 years and were freshmen, juniors and seniors. Half of the participants had spent time outside of Palestine, generally for periods of less than 6 months in either European or Arab countries. Only one had spent time in the US. All had learned their English primarily in Palestine.

Table 2. Characteristics of student participants in Study 1

Characteristic	Institution 1 (n=15)	Institution 2 (n=15)	Total (n=30)
Gender			
Male	8 (53.3%)	7 (46.6%)	15 (50.0%)
Female	7 (46.6%)	8 (53.3%)	15 (50.0%)
Year of Study			
Freshman	6 (33.3%)	0 (0.0%)	6 (20.0%)
Junior	9 (66.7%)	4 (26.7%)	13 (43.3%)
Sophomore	0	7 (46.7%)	7 (23.3%)
Senior	0	4 (26.7%)	4 (13.3%)
Mean English Score	B+ (87-90%)	B (83-87%)	
Time spent outside Palestine			
Less than 6 months	8	2	10 (33.3%)
6-12 months	0	0	0 (0.0%)
13-24 months	1	0	1 (3.3%)
25-36 months	1	0	0 (0.0%)
More than 36 months	2	2	4 (13.3%)

3.1.2 Testing Procedures

Students completed the CAT test at their institution in a quiet room and were supervised by the first author. Students were not given any course credit or monetary incentive for completing the test. The CAT is not a timed test, so students were told that they could take as much time as they needed. The completed tests were scored by a team of experienced scorers who had been trained in how to score the CAT test by the designers of the test.

Immediately after they finished the test, all 30 students completed a survey in English asking them if there were any aspects of the test directions or content that were confusing. They were also asked to rate the difficulty of the test on a scale from 1 (very difficult) to 7 (very easy) and their interest in the test again on a 7-point scale from 1 (very interesting) to 7 (not very interesting at all). Students were invited to include explanations for their answers. The first author also noted comments that students made to him directly after completing the test.

3.2 Results

All 30 students completed the test. Time taken to complete the test ranged from 1 to 2 hours. Scores on the test ranged from 5 to 27 (out of a possible 38) with a mean of 16.4 and standard deviation of 5.7. The mean score fell above the US norm for community colleges (13.5) and between the US norms for freshmen at 4 year institutions (13.7) and seniors at 4 year institutions (19.0). Overall, the students responded favorably to the test, reporting that they found it interesting and motivating. Even though the test questions were developed to fit the cultural context in the US, only 2 students reported on the survey that the content was problematic. One explained the following; "As the questions are related to cases in foreign country, it is difficult to think for possible answers." A second wrote that "Because it was my very first time reading about purification". The same student suggested that it would be easier if the questions were related to cases in their own country. One third (n=10) of the students reported that the test directions were confusing. However, when asked to explain what was confusing, most did not describe a problem understanding the directions, but rather referred to challenges with the type of thinking that was required. For example, one student wrote "When I know that it is a critical thinking assessment, I started criticizing everything and said no to almost every question. May be I should not have been told to give more accurate information". Another student wrote "It is complicated. Too many answers needed to be written with explanations". A third student wrote "It depends on my analytical competencies". Only one student made reference to a specific aspect of the test instructions that was confusing. "The moment you were told the type of study which can include a third variable, the last question was quite confusing."

Seven of the 30 students found some of the test questions confusing. This seemed to be mainly because some of the English vocabulary was new to them. Only 2 students commented on the actual content of the test. A second student reported that it was their first time reading about water purification.

Students varied in their opinions about the difficulty of the test. Some found the test very hard, some very easy with the majority finding it moderately difficult. Students' ratings of the difficulty of the test ranged from 2 to 7 (where 1 is very difficult and 7 is very hard). Students also varied in how well they thought they did on the test. Most felt that had done moderately well to well. This is consistent with the responses of college students who have taken the test in the US (Stein, 2012). There was a high level of interest in the test which is also consistent with findings in the US (Stein et al., 2009) with a number of students reporting either that it was a new experience for them or a new way of thinking, and several commenting that they enjoyed the test and the challenge of taking it. Interest ratings ranged from 1 to 7 (where 1 is very interesting and 7 is not very interesting at all), with two thirds of the sample choosing a rating of either 1 or 2). Comments collected from the students after they finished the test shed some light on their capacity to deal with the CAT test. The majority of students believed that they had made substantial gains in their ability to think critically after completing the test. Some students asked to receive instruction in critical thinking, commenting that they had never been exposed to such challenging questions. Some students said that because the test was new for them that they were concerned that they might do the wrong thing and this was stressful for them. At the end of the test a number of students said that they were tired because of the mental effort that the test required.

The mean total score for males was 16.4 (sd=6.0) and the mean total for females was 16.5 (sd=5.5). An independent t-test revealed that this difference was not statistically significant, suggesting no gender bias in the test.

3.3 Discussion

Contemporary Palestinian higher education is very different from the modern university environment which prevails in western countries, especially from the environment which exists in the US. In addition to its Middle Eastern location and distinctive cultural character, today's Palestinian university perseveres and flourishes in a unique socio-political context. Characterized from the beginning by occupation and conflict (Abu-Lughod, 2000; Abu-Saad & Champagne, 2006), university students in Palestine live and study today in a social context fundamentally unrecognizable to present day American students. Nevertheless, despite these considerable differences, the results of the study above suggests that the CAT test developed in the U.S. is a valid and meaningful instrument for assessing the critical thinking skills of Palestinian university students, at least of those students whose English language ability is sufficient to complete the English version of the test. Quantitative test results reflected similar results to those obtained with students in American colleges and universities at the higher education level. Palestinian students scored within the range that US students fall and there was no statistically significant difference between the total scores of males and females. These results suggest that the test may not be culturally or gender biased. Qualitative responses of their experience of the test, moreover, reflected the experience of American students. Although one third of the sample who took the test in English reported that the directions for the test were confusing, their confusion seemed to relate chiefly to the challenging nature of what they were being asked to do, rather than to confusing language. Their comments echoed those of many American students who noted their traditional American education did not stress the kinds of thinking skills which the CAT assesses. The essential skills which the CAT assesses were no more alien to or difficult for Palestinian students than they were for American students. Indeed the results suggest the test gauges higher order thinking skills which are common to both sets of students; skills which are, moreover, sought after by educational policy commentators in both contexts (Arum & Roksa, 2011; Fasheh, 2014; AbuLaban, 2014).

It is worth noting that the Palestinian students in the above study were all fluent in English and many had spent time abroad. These students did not necessarily reflect the typical Palestinian University student. Their skills with a second language-particularly English-and experiences abroad might explain the similarity of their experience to those of American students. Additional study of Palestinian students taking the test in Arabic is warranted and we do this in Study 2 below.

4. Study 2: Students' Responses to the First 4 CAT Questions Translated into Arabic

The aim of this study was to assess Palestinian students' responses to the first 4 questions of the CAT test which had been translated into Arabic.

4.1 Methods

4.1.1 Participants

Forty-eight students from the same 3 Palestinian universities as the students in study 1 participated in study 2. Fifty percent were female and 50% were male. Twenty-four were freshman and 24 were seniors. Students from the faculties of medicine, education and business were invited to participate in the study by their course professors. The first author contacted a number of faculty at each institution and asked if they would nominate students who might be interested in answering the 4 questions. A few instructors asked if they should nominate their best students. They were told just to invite those students who they thought would be interested in doing the test. Students were told that the 4 questions were part of a longer test of critical thinking. They did not receive any incentive, course credit or other reward for answering the questions.

4.1.2 Testing Procedures

The first 4 questions of the CAT test were translated into Arabic by the first author. The first 4 questions focus on critical thinking skills 1 to 4 listed in Table 1. Translation of the 4 questions was validated by 4 Palestinian university faculty before they were given to students. The first author administered the test questions, and was present while students answered them. Once again, students completed the questions in a room under quiet conditions. They were told that their responses would be anonymous and that no identifying information would be collected. They were given about 45 minutes to answer the 4 questions which was considered sufficient time as the majority of students in the US complete the entire test in just under 1 hour. Participants did not complete a formal survey after answering the 4 questions. However, the first author was present during and after testing and collected comments from participants after they had completed the test.

4.2 Results

All 48 students completed the Arabic version of the first 4 questions. Most of the students reported that they had an easy time answering the questions. A few complained about having not been administered a similar test in the past. Some students seemed nervous while answering the questions. A couple of students asked to have more time to answer the questions while a few others stopped suddenly during the test and asked if they were allowed to leave the hall to come back after having a quick break. They were not given a break however. Many of the students offered favorable comments about this opportunity to take the test and said that they wished that their own teachers would focus on this.

Mean scores on the 4 items for the 48 students are presented in Table 3 along with those of the students who took the English version of the CAT, and the norms for US college freshmen and seniors. Mean scores for the Palestinian students were above those of US students for 2 items and within the range of freshmen and seniors for one item. Only scores for the first item fell below those of US students. It is interesting to note the higher scores of the Palestinian students on items 3 and 4. This may have been because a number of the students were medical students who were familiar with critical thinking that is often applied in observational studies in public health.

Table 3. Mean scores on the first 4 questions on the Arabic and English versions of the CAT compared with norms for US freshmen and seniors

Question	Palestinian students Arabic (n=48)	Palestinian students English (n=27)	US college student norm for Freshmen*	US college student norm for seniors*
Q1 (1 point max.)	0.46	0.74	0.58	0.67
Q2 (3 points max.)	0.96	1.52	0.69	1.21
Q3 (3 points max.)	1.64	1.0	0.67	1.35
Q4 (4 points max.)	1.71	0.85	0.96	1.41

Q1: Summarize the pattern of results in a graph without making inappropriate inferences.

Q2: Evaluate how strongly correlational-type data supports a hypothesis.

Q3: Provide alternative explanations for a pattern of results that has many possible causes.

Q4: Identify additional information needed to evaluate a hypothesis or a particular explanation of an observation.

4.3 Discussion

To ascertain whether the results of the first study would hold when the test was not given to Palestinian students fluent in English, the second study investigated the experience of Palestinian students not fluent in English taking the CAT test in Arabic (Hambleton, 2005). While this follow up study only focused on the critical thinking skills measured by the first four questions of the CAT test, the results of the second study support the premise that the critical thinking skills assessed by the CAT are appropriate to Palestinian higher education students more broadly and that an Arabic adaptation of the CAT might be successfully developed and employed to assess the critical thinking skills of Palestinian students in their main language of instruction. While the results suggest the Arabic items of the test are meaningful to the students, differences on some items such as item 1 indicate more refinement may be required for particular items and certainly in a full adaptation of the entire test (Sireci, 2005). The results do indicate a full adaptation of the CAT in Arabic would be valuable.

In addition, it is worth noting that the comments (and symptoms of some anxiety) from students about their lack of preparation for the kind of thinking the test measures, and their interest in having instruction in this kind of thinking, highlight the importance of the role of the instructor in the use and application of such a test. Engagement of the faculty with the key critical thinking skills is necessary if they are to develop further in Palestinian universities.

5. Study 3: Faculty Perceptions of Critical Thinking Skills Assessed on the CAT

The aim of the third study was to assess the face validity of the CAT for Palestinian faculty. For a critical thinking assessment to be useful in any context, it is important that faculty agree and have confidence that it is a measure of critical thinking. This is particularly important in the case of the CAT because it was designed to be used not only for assessment, but also for faculty development. Although there is a high level of agreement among STEM faculty in the US that the skills tested on the CAT constitute critical thinking (Stein, Haynes, & Redding, 2006), it is important to establish whether or not this is the case for Palestinian faculty.

5.1 Methods

5.1.1 Participants

All Palestinian universities and university colleges in both the Palestine were invited to participate in the study. The first author phoned the Vice President for Academic Affairs at each institution to gain their approval for faculty at their institution to participate in the study and followed the telephone call up with an official letter of invitation to participate in the study. In addition, the first author reached out to individual faculty at the universities and colleges. The study was described as looking at one aspect of educational skills at Palestinian universities from the perspective of both students and faculty.

One-hundred-twelve faculty responded to the survey. Characteristics of respondents are summarized in Table 4. The majority of participants were full-time male faculty. This is representative of faculty at Palestinian Universities. Forty-two percent held PhD's which is also representative of Palestinian university faculty. A variety of disciplines were represented including business, the humanities, medicine and science math and engineering. The majority of the sample had at least 2 years of teaching experience, with approximately one third having more than 10 years of teaching experience. It should be noted that there was quite a bit of missing data for both demographic questions and questions about critical thinking. Missing data rates ranged from 14% to 15% for demographic questions and 17% to 35% for the critical thinking skills. We are unsure of the reason for this relatively high rate of missing data. It may have been due to faculty not being used to this type of survey or to concern about revealing data which they may have felt was identifiable. In the case of the questions about critical thinking skills, participants may have fatigued during the survey and/or may not have fully understood the critical thinking skills. Also, as there were no incentives for completing the survey, faculty may have lacked motivation.

Table 4. Characteristics of faculty who participated in the study

Characteristic	Number of Participants (%) N=112
Gender	
male	71 (63.4)
female	25 (22.3)
missing	16 (14.3)
Position	
full-time faculty	67 (75.0)
part-time faculty	11 (9.8)
administrator	7 (6.3)
other	5 (4.5)
missing	22 (19.6)
Discipline	
science, math engineering or technology	30 (26.8)
humanities	24 (21.4)
other	15 (13.4)
medical	13 (11.6)
missing	18 (16.1)
Education	
PhD	47 (42.0)
MA	38 (33.9)
missing	27 (24.1)
Years of teaching experience	
0-1	10 (8.9)
2-5	15 (13.4)
6-10	19 (17.0)
>10	42 (37.5)
missing	26 (23.2)
Age	
<34	20 (17.9)
35-44	23 (20.5)
45-54	26 (23.2)
>54	15 (13.4)
missing	28 (25.0)
Province	
West Bank	78 (69.6)
Jerusalem	10 (8.9)
Gaza	8 (7.1)
missing	16 (14.3)

5.1.2 Procedures

Participants completed an on-line survey. Questions were in both Arabic and English and participants were free to respond in which ever language they were most comfortable with. The first section of the survey asked participants for demographic information. In the second section, participants were presented with the 15 skills tested on the CAT. For each skill, they were asked to indicate whether they agreed that the skill was a dimension of critical thinking or not. They were also asked whether they felt the skill was important and relevant to their teaching and a number of other questions about their understanding and teaching of critical thinking. We report here just the demographic data and data on agreement/disagreement with the critical thinking skills. Data on relevance of the skills to their teaching and other survey items will be reported in an upcoming paper.

5.2 Results

Overall there was a moderate level of agreement that the skills tested on the CAT represent critical thinking. Levels of agreement ranged from 42.0% for "Evaluate whether spurious relationships strongly support a claim" to 73.2% for "Evaluate how strongly information supports a hypothesis or interpretation". These rates are lower than those reported with faculty in the US where agreement was at least 80% across all the skills (Stein, Haynes, Redding, Ennis, & Cecil, 2007). However, this may have been because the US sample included only STEM faculty, and because the Palestinian sample included both STEM and non-STEM faculty. When agreement was examined in STEM faculty only, the agreement rate was much higher and was 80% or above for all but three skills (Table 6). The percent agreeing may also be lower in the Palestinian sample because data were missing for between 17% and 35% of the sample. This was higher than the missing data rate for the demographic data which ranged from 13% to 25%. This may have been due to difficulty comprehending and interpreting the CAT skills or time constraints/survey fatigue which prevented busy faculty from completing the survey.

Table 5. Percent of respondents who agreed and disagreed that skills tested on the CAT test represent dimensions of critical thinking

CAT Question	Skill	Percent Yes	Percent No	Did not respond
1	Summarize a pattern of information without making inappropriate inferences	52.7	29.4	17.8
2	Evaluate how strongly information supports a hypothesis or interpretation	73.2	4.4	22.3
3	Provide alternative explanations for observations	68.7	8.0	23.2
4	Identify additional information needed to evaluate a hypothesis or particular explanation of an observation.	61.6	8.9	29.5
5	Evaluate whether spurious relationships strongly support a claim	42.0	27.7	30.3
6	Provide alternative explanations for spurious relationships	50.0	19.6	30.3
7	Identify additional information needed to evaluate a hypothesis/interpretation	59.8	9.8	30.3
8	Determine whether an invited inference in an advertisement is supported by information	48.2	17.8	33.9
9	Provide relevant alternative interpretations of information	59.8	4.9	34.8
10	Separate relevant from irrelevant information when searching for information to solve a real-world problem	52.7	14.3	33.3
11	Analyze and integrate information from separate sources to solve a real-world problem	59.8	4.5	35.7
12	Use basic mathematical skills to help solve a real-world problem	47.3	16.9	35.7

13	Identify suitable solutions for a real-world problem using relevant information	59.8	4.5	35.7
14	Identify and explain the best solution for a real-world problem using relevant information	58.9	5.3	35.7
15	Explain how changes in a real-world problem situation might affect the solution	60.7	6.3	33.0

Table 6. Percentage of STEM and non-STEM faculty who agreed that skills tested on the CAT are dimensions of critical thinking

CAT Question	Skill	Percent STEM Yes (n=30)	Percent Non-STEM Yes (n=52)
1	Summarize a pattern of information without making inappropriate inferences	70.0	57.7
2	Evaluate how strongly information supports a hypothesis or interpretation	100	78.8
3	Provide alternative explanations for observations	93.3	75.0
4	Identify additional information needed to evaluate a hypothesis or particular explanation of an observation	83.3	67.3
5	Evaluate whether spurious relationships strongly support a claim	46.6	50.0
6	Provide alternative explanations for spurious relationships	60.0	59.6
7	Identify additional information needed to evaluate a hypothesis/interpretation	83.3	61.5
8	Determine whether an invited inference in an advertisement is supported by information	60.0	55.8
9	Provide relevant alternative interpretations of information	86.7	65.2
10	Separate relevant from irrelevant information when searching for information to solve a real-world problem	80.0	53.8
11	Analyze and integrate information from separate sources to solve a real-world problem	93.3	59.6
12	Use basic mathematical skills to help solve a real-world problem	70.0	40.4
13	Identify suitable solutions for a real-world problem using relevant information	90.0	51.9
14	Identify and explain the best solution for a real-world problem using relevant information	86.7	61.5
15	Explain how changes in a real-world problem situation might affect the solution	86.7	65.4

5.3 Discussion

Faculty recognition of the validity of the critical thinking assessment test is essential for both its acceptance and implementation in Palestinian institutions of higher education. Even more importantly, perhaps, recognition of such skills is vital for instructional changes to be made in the teaching and learning environment (Light, Cox, & Calkins, 2009) to achieve the critical thinking skills being advocated for nationally. The face validity of the test for faculty is a critical condition of both its use and its potential for instructional change. In this respect, the third

study revealed moderate to high agreement among Palestinian faculty that the skills tested on the CAT are key dimensions of critical thinking, with even greater agreement among STEM faculty. This suggests that the test has a reasonable degree of face validity. The results are especially remarkable among the STEM faculty. This is of particular significance given the weight which the Palestinian educational strategy places on the development of critical thinking skills for employability and their focus on science and technology in their work and employment strategies (Palestinian National Authority, 2012).

6. Conclusions

The results of the 3 studies suggest that the CAT test has potential as an assessment tool for critical thinking in Palestinian higher education. It could be particularly useful as part of an national quality assessment strategy for improving learning (often missing in broader policy debates), and which has been recommended in a recent national report on undergraduate teaching practices in Palestinian higher education (Cristillo, 2009) and by the recently established Association of Palestinian Academic Developers. Together these results suggest that a large scale validation study of the CAT test in Arabic would be worthwhile. We recommend that CAT be carefully translated into Arabic—in accordance with international guidelines (Hambleton, 2005)—and tested on a large representative sample of students across Palestinian Universities. We also recommend that a full Differential Item Functioning (DIF) analysis be conducted to determine if any components of the test are biased towards certain demographic groups (Schmeiser & Welch, 2006). In addition to the larger study with students, we also recommend that a more detailed study on faculty attitudes towards the CAT be conducted. This would involve having faculty examine the test directly, receive training in how to score the test, and participate in a test scoring session. This step is particularly important because if the CAT is to be utilized for assessing critical thinking outcomes in Palestinian students it must have face validity for faculty and be considered as both practical and meaningful.

There is, moreover, a great deal of interest in developing critical thinking in students across the Middle East (Brewer et al., 2006; Al-Essa, 2009; Romanowski & Nassar, 2012). Therefore, an Arabic version of the CAT test is likely to have value well beyond Palestine as a means of measuring progress towards critical thinking goals at national, institutional and individual student levels.

Acknowledgements

References

AbuLaban, R. (2014). Why do we pursue university degrees? *Palestine This Week*, *199*(11), 28-33.

Abu-Saad, I., & Champagne, D. (2006). Introduction: A Historical Context of Palestinian Arab Education. *American Behavioral Scientist*, *49*, 1035. http://dx.doi.org/10.1177/0002764205284717

Accreditation and Quality Assurance Commission. (2010). *CINDA Manual E./.* Ramallah, West Bank.

Al-Essa, A. (2009). *Education Reform in Saudi Arabia between Absence of Political Vision, Apprehension of the Religious Culture and Disability of Educational Management* (1st ed.). Beirut: Dar Al Saqi.

AmidEast. (PDFP). Retrieved from http://www.amideast.org/pfdp/program-components/about-pfdp

Arum, R., & Roksa, J. (2011). *Academically Adrift: Limited Learning on College Campuses*. University of Chicago Press, Chicago and London.

Basha, S. (2012). *Special and inclusive education specialization in mild and moderate disabilities*. EDUCAT, Milan.

Bernhard, A. (2012). Quality Assurance in an International Higher Education Area: A summary of a case-study approach and comparative analysis. *Tertiary Education and Management*, *18*(2), 153-169. http://dx.doi.org/10.1080/13583883.2012.654504

Bloom, B. S. (1956). *Taxonomy of Educational Objectives*. Longmans Green, NY: New York.

Brewer, D., Augustine, C., Zellman, G., Ryan, G., Goldman, C., Stasz, C., & Constant, L. (2006). *Education for a new era design and implementation of K-12 education reform in Qatar*. Rand Qatar Policy Institute, Santa Monica, CA: Rand Corporation.

Bruhn, C. (2006). Higher Education as Empowerment: The Case of Palestinian Universities. *American Behavioral Scientist*, *49*(4), 1125-1142. http://dx.doi.org/10.1177/0002764205284722

Cristillo, L. (2009). *National study of undergraduate teaching practices in Palestine*. Palestinian Faculty Development Program, AMIDEAST, Ramallah, West Bank.

Daragmeh, A. K., Drane, D., & Light, G. (2012). Needs assessment and beyond in the setup of centers for teaching and learning excellence: An-Najah university center as a case study. *Procedia-Social and Behavioral Sciences*, *47*, 841-847. http://dx.doi.org/10.1016/j.sbspro.2012.06.745

Fasheh, M. (2014). Healing from modern superstitions. *Palestine This week*, *199*(11), 10-19.

Hambleton, R. (2005). Issues, designs and technical guidelines for adapting tests into multiple languages and cultures. In *Adapting Educational and Psychological Tests for Cross Cultural Assessment*. New Jersey: Lawrence Erlbaum Associates.

Jabr, D. (2003). Teaching critical thinking in the English language classroom: The case of Palestine. *Mediterranean Journal of Educational Studies*, *8*(2), 137-162.

Kassab, E. S. (2009). Contemporary Arab Thought: Cultural critique in comparative perspective. In *Digest of Middle East Studies* (Vol. 20, No. 1). Columbia University Press, New York, NY.

Light, G., Cox, R., & Calkins, S. (2009). *Learning and Teaching in Higher Education*. Los Angeles, USA: Sage.

Lughod, I. (2000). Palestinian Higher Education: National Identity, Liberation and Globalization. *Boundary 2*, *27*(1), 75. http://dx.doi.org/10.1215/01903659-27-1-75

MINEDUC. (2010). *Education Sector Strategic Plan 2010-2015*. Kigali, Rawanda. Republic of Rwanda Ministry of Education.

Palestinian National Authority. (2012). *The National Strategy to Achieve the MDGs by 2015*. Ramallah, West Bank: Palestinian National Authority.

Rauhvargers, A. (2011). *Global University Rankings and their Impact*. *A report for the European Universities Association (EUA)*. Brussels, EUA.

Romanowski, M. H., & Nassar, R. (2012). Critical Thinking in Qatar's education for a new era: Negotiating possibilities. *International Journal of* Critical *Pedagogy*, *4*(1), 118-134.

Schendel, R. (2013). *A critical missing element: Critical thinking at Rwanda's public universities and the implications for higher education reform* (Unpublished doctoral dissertation). Institute of Education, University of London, London, U.K.

Schendel, R. (2015). Critical thinking at Rwanda's public universities: Emerging evidence of a crucial development priority. *International Journal of Educational Development*, *42*, 96-105. http://dx.doi.org/10.1016/j.ijedudev.2015.04.003

Schmeiser, C. B., & Welch, C. J. (2006). Test development. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed.). Westport, CT: American Council on Education and Praeger Publishers.

Shonstrome, E. (2014). People Over Pedagogy. *Chronicle of Higher Education*, *60*(43), B4-B5.

Sireci, S. (2005). Using bilinguals to evaluate the comparability of different language versions of a test. In *Adapting Educational and Psychological Tests for Cross Cultural Assessment*. Mahwah New Jersey and London: Lawrence Erlbaum Associates,

Stein, B., Haynes, A., & Redding, M. (2006). Project CAT: Assessing critical thinking skills. *Proceedings of the National STEM Assessment Conference*. Hosted by the National Science Foundation and Drury University.

Stein, B., Haynes, A., Redding, M., Ennis, T., & Cecil, M. (2007). Assessing critical thinking in STEM and beyond. In M. Iskander (Ed.), *E-learning, Instruction Technology, Assessment and Engineering Education* (pp. 79-82). http://dx.doi.org/10.1007/978-1-4020-6262-9_14

Stein, B., Haynes, A., & Harris, K. (2009). *Assessing and Improving Critical Thinking Using the CAT Instrument*. ABET Best Assessment Processes Symposium XI. Indianapolis, IN: Accreditation Board of Engineering and Technology.

Stein, B., & Haynes, A. (2011). Engaging Faculty in the Assessment and Improvement of Students' Critical Thinking Using the Critical Thinking Assessment Test. *Change: The Magazine of Higher Education*, *43*(2), 44-49. http://dx.doi.org/10.1080/00091383.2011.550254

Stein, B., Haynes, A., & Drane, D. (2012). *Getting faculty involved in assessing and improving students' critical thinking*. Presentation at the Annual Conference of the Higher Learning Commission, April 15-19. Chicago, USA.

U.S. Department of Education, National Center for Education Statistics. (2000). The NPEC Sourcebook on Assessment, Volume 1: Definitions and Assessment Methods for Critical Thinking, Problem Solving, and Writing, NCES 2000—172. In T. D. Erwin (Ed.), *For the Council of the National Postsecondary Education Cooperative Student Outcomes Pilot Working Group: Cognitive and Intellectual Development*. Washington, D.C.: U.S. Government Printing Office.

Zellman, G. L., Martini, J., & Perlman, M. (2011). *Identifying Arabic-Language Materials for Children That Promote Tolerance and Critical Thinking*. National Defense Research Institute.

Note

Note 1. The Palestinian Faculty Development Project (PFDP) was an 8 year project funded by USAID and the Open Society and implemented by AmidEast. Teaching Centers were established at An-Najah University, Bethlehem University, the Palestinian Polytechnical University (PPU) and the Palestinian Technical University-Khadoori (PTU-K). A description of the project can be found at http://amideast.org/pfdp/program-components/about-pfdp