International Journal of Teaching and Learning in Higher Education 2019, Volume 31 Number 3, 402-412
http://www.isetl.org/ijtlhe/ ISSN 1812-9129

Explicit Conditional Reasoning Performance: An Examination of Educational Past, Processing Load, and Self-Efficacy

Maura A. E. Pilotti, Siddiqua Aamir, Runna Al Ghazo, and Halah Abdulaziz Al Kuhayli
Prince Mohmmad Bin Fahd University

The main aim of the present investigation was to examine conditional reasoning skills in college students whose educational past had emphasized verbatim learning. A successive independent-samples design was utilized to explore the effects of instruction that explicitly targeted critical thinking principles in either freshman students or sophomores. Conditional reasoning scores of freshman students were not higher than those of sophomores, even when the impact of either GPA or self-efficacy was statistically controlled. Furthermore, the students in our sample performed as well as students with a similar educational past, whereas both scored below students whose education had deemphasized verbatim learning. In addition to past educational practices, differences in performance arose from processing load. Not surprisingly, self-efficacy and processing load (as determined by a test read in the second language), but not GPA, predicted conditional reasoning scores. We conclude that demanding cognitive computations, such as those of a conditional reasoning test taken in a second language, not only reflect the test-taker's knowledge, but also are sensitive to processing load, and past educational practices, as well as self-efficacy since confidence in one's abilities translates into effort and persistence.

It has been said, perhaps too many times, that one of the striking characteristics of formal education in the Arab world is its reliance on rote learning, including memorization and recitation (Rugh, 2002). This pedagogy emerges from the oral tradition of early communities for whom memorization and recitation were means to preserve scriptures and remember the past, as well as activities contributing to knowledge acquisition, understanding, self-discipline, and reasoning (Douglass & Shaikh, 2004). Since learners are envisioned as passive knowledge recipients, even in problem-solving situations, they are expected to retain, rather than generate, answers to fairly fixed questions. As a result, calls to develop instructional practices whose goal is to nurture critical thinking capacities have become louder, but have not translated into unequivocally effective interventions (see Lehman & Nisbett, 1990; Tirunch, Verburgh, & Elen, 2014). Not surprisingly, different viewpoints have emerged not only about the most effective form of instruction, but also about the mere definition of critical thinking above and beyond its generic characterization as "reasonable and reflective thinking focused on what to believe or do" (Ennis, 2011; p. 1). Namely, the term critical thinking is used to refer to "good strategies" (Nickerson, Perkins, & Smith, 1985), cognitive skills and dispositions which are conducive to effective decision making and problem solving in different situations (Ennis, 1987; Halpern, 1998).

Rationale of the present Investigation

There are different types of critical thinking (Ennis, 1964; Pascarella & Terenzini, 2005). In the present investigation, we focus on conditional reasoning, a form of information processing requiring that a conclusion be drawn from premises. *Modus ponens* is an example of a conditional argument that possesses two related premises of the form, "If p, then q," and "p" (p is the antecedent and q is the consequent), as well as a conclusion of the form "Therefore, q".

In essence, conditional reasoning entails drawing inferences (conclusions) about situations in which the occurrence of one event is conditional or contingent upon the occurrence or non-occurrence of another event (premises). It is thus an essential mode of thinking in daily life whose study has mostly focused on cognitive and attentional factors as the primary sources of individual differences (Barrouillet & Lecas, 1999; Cummins, Lubart, Alksnis, & Rist, 1991). The present research takes a slightly different approach. It begins with the recognition that the nurturing of critical thinking, including conditional reasoning, is a prominent goal of university education (see Pithers & Soden, 2000) and that good strategies can be taught and measured objectively. It is a field research that examines the extent to which knowledge of principles of conditional reasoning, explicitly taught in a course, is possessed by a particular kind of students at two points of the undergraduate curriculum (freshman and sophomore years). Its targets are college students whose past educational experiences have put a premium on rote learning. In college, these students are asked to adapt to a mode of instruction that includes analysis, inference, evaluation, explanation, and interpretation of information and that relies heavily on self-regulation. A standard test of conditional reasoning, the Cornell Conditional-Reasoning, CCR, test-Form X (Ennis et al., 1964), which assesses students' formal knowledge of conditional reasoning principles, is used to measure the extent to which students have interiorized this initially foreign mode of information processing. The argument

that human information processing is shaped by the experiences of one's society and culture—which may create habits, dispositions, and skills—is not novel. It has propelled accounts of test performance differences between Westerners and other cultural/social collectives, such as East Asians (Nisbett, Peng, Choi, & Norenzayan, 2001), albeit evidence has not always been supportive (Lun, Fischer, & Ward, 2010). The test is written in the students' second language (English), which places an additional burden on the cognitive resources of the test-takers due to their concurrent use of a mode of processing yet to become a habit. The questions and the ensuing predictions that guide our investigation are as follows:

A. Will freshmen's knowledge of conditional reasoning principles, which was explicitly taught in a general education class, differ from that of sophomores? To answer this question, the CCR test-Form X (Ennis et al., 1964) is used to assess the status of students' formal knowledge of conditional reasoning principles (i.e., stability, improvement, or decline) at two points of the university curriculum (i.e., freshman and sophomore years). A successive independent-samples design with two groups is used: freshmen who are about to complete a course where conditional reasoning principles have been formally taught (formal instruction condition serving as baseline) and sophomores who have completed the class approximately a year earlier (post-instruction condition to measure status of possessed knowledge as a function of the passage of time). Performance feedback (e.g., class discussion of test performance) is likely to generate carry-over effects on re-testing, thereby preventing the use of a longitudinal design. Because the successive independent-samples design permits feedback to closely follow the test-taking experience, it is selected to ensure that the experience of taking a conditional reasoning test is educational rather than merely an opportunity for research. It is predicted that since most university classes taken concurrently and after the baseline class tend to emphasize the relevance of critical thinking (defined as "reasonable and reflective thinking focused on deciding what to believe and do"; (Ennis, 2011; p. 1) in everyday life and explore generic applications, knowledge of conditional reasoning principles may remain active (see Nisbett, 2013) and even be improved by practice. Alternatively, knowledge of conditional reasoning principles may degrade if practice does not draw attention to the link between applications and formal knowledge (Ausubel, 2012; Nisbett, Fong, Lehman, & Cheng, 1987).

B. Is the conditional reasoning performance of students whose past educational experiences emphasize rote learning different from that of students who underwent an education deemphasizing such learning? Rote learning is the main feature of an instructor-centered education, whereby instruction and instructional resources are not to be contradicted or criticized because they are the artifacts of experts whose job description is to impart knowledge (Oettingen, 1995; Stipek, 1991). Students' ability to reiterate study and lecture materials verbatim is often mistakenly assumed to be a sign of mastery. Verbatim learning becomes a disposition that has been shaped and reinforced by a pedagogy, widespread in schools of the Arab world, which emphasizes the practice of memorization and recitation (Douglass & Shaikh, 2004). This practice promotes rapid, but short-lived acquisition of knowledge and is largely inadequate to the demands of learning in college which entail, in addition to remembering and understanding, application, analysis, evaluation, and generation of knowledge. In contrast, one of the main features of a learner-centered education is that learning depends mostly on personal choices and is an active endeavor whereby knowledge can be manipulated to address issues, solve problems, and create solutions (Oettingen, 1995; Stipek, 1991).

For undergraduate students whose past scholastic experiences have been shaped by an instructor-centered education, practical knowledge of test and class demands may be an especially potent agent of change. Moreover, explicit conditional reasoning instruction may become a welcome opportunity to develop and practice reasoning skills whose utility encompasses many of the classes students take to complete their degree. On the other hand, resistance to change may also be a potent force even in the face of unavoidable class demands. Thus, the breadth of the impact of explicit conditional reasoning instruction may not go beyond the course taken. The existing literature is vague as to the impact of such instruction, particularly in the case of recipients whose educational background is instructor-centered (Al-Ghamdi & Deraney, 2013; Al-Wehaibi, 2012; Tirunch et al., 2014). For instance, in a study using a longitudinal design, normatively poor pre- and post-test performance was reported for students from the Kingdom of Saudi Arabia (KSA), even though moderate gains were detected at post-test (Al-Ghamdi & Deraney, 2013), whereas in a comparable study, gains were observed (Al-Wehaibi, 2012).

C. What are some of the factors that account for knowledge of conditional reasoning principles as measured by the CCR Test-Form X (Ennis et al., 1964) in students with an instructor-centered educational past? We consider self-efficacy (illustrating confidence in one's abilities), GPA (serving as a generic index of effort, persistence, and capabilities), and processing load.

Evidence exists that general self-efficacy, an optimistic sense of personal competence, is positively correlated with task completion rates (Eden, 1984, 1988;

Pajares, 1996), persistence (Bandura, 1977; Pajares, 1997), and motivation and engagement (Bandura, 1989; Bandura & Schunk, 1981). Tests of conditional reasoning, such as the CCR TestForm X (Ennis et al., 1964), are demonstrably challenging for college students (see McLellan, 2012). Thus, the allocation of cognitive resources to organize and energize challenging computations to solve conditional reasoning riddles is expected to reflect students' confidence in their competence (Pajares, 1996). A similar prediction is made for GPA, treated as a rough index of achievement motivation. However, evidence regarding the correlation between GPA and conditional reasoning performance is mixed. For instance, McLellan (2012), and Lehman and Nisbett (1990) failed to find a correlation, whereas Johnson and Posner (1971) reported a moderate one.

Important to note though is that a conditional reasoning test, such as the CCR test-Form X (Ennis et al., 1964), not only may demand a great deal of computational resources, but also is written in English. Evidence exists that performance is better if an assessment tool is written in the first language of the test-takers (Campbell, Adams, & Davis, 2007; Campbell, Dollaghan, Needleman, & Janosky, 1997) and that the lower performance of second-language test-takers may be related to the cognitive demands of second language processing as well as to the culturally biased content of test items, making reading comprehension more effortful or even problematic (Hambleton, Merenda, & Spielberger, 2004). Thus, it is not surprising that students who have English as their second language perform less well on the CCR test-Form X (Ennis et al., 1964) than native English speakers (McLellan, 2012; Nolan & Brandon, 1984), as second language processing may place a burden on an already overloaded cognitive system (Campbell et al., 2007; Paas, Renkl, & Sweller, 2003). If the sustained processing load, which arises from reading-comprehension of a test written in the second language of the test-takers (Takano & Noda, 1993), unfairly depresses performance, using items that assess a reduced number of conditional reasoning principles, rephrasing a few culturally opaque items or even translating the test in the first language of the test-takers may counteract such an effect. Yet, a student's educational past may be the critical factor that curbs the benefits of processing load reductions.

Questions a-c and corresponding predictions define the scope of our investigation. The methodology described below illustrates how predictions were tested.

Method of Study 1

Participants

The participants were 467 undergraduate students from a private university located in the Eastern Province of KSA. They were Arabic-English bilingual speakers whose age ranged from 18 to 25. For university admission, students had demonstrated English language competence through standardized English proficiency tests (i.e., IELTS, Aptis, or TOEFL). Answers to queries based on Weimer's dimensions differentiating educational approaches (2002) were used to classify participants as possessing an instructor-centered educational past.

Procedure and Materials

Students were enrolled in one of two mandatory, sequentially arranged classes of the general education curriculum: critical thinking (n = 111) and learning outcome assessment (n = 356). Critical thinking is a course that explicitly teaches freshmen the formal principles of reasoning and offers practice in their application to real-life contexts. Instead, assessment, usually taken by sophomores, entails a review of the general properties of reasoning (e.g., clarity, precision, accuracy, relevance, significance, completeness, logicalness, fairness, depth, and breadth; Paul & Elder, 2014) as they apply to self-assessment. Although references to conditional reasoning are interspersed across the entire critical thinking course, one of the four units explicitly focuses instruction on conditional reasoning principles and fallacies. The two courses are completed either a semester or two apart, depending on the academic program. Since no effects of academic program or time separating courses were found in the analyses described below, this factor was not considered further. Through convenience sampling, four sections of critical thinking out of 6 (67% of the available classes) and 18 sections of assessment out of 27 were selected (67% of the available classes) during the fall semester. Sampling relied on assent of the instructor and equitable distribution of morning and afternoon classes.

Towards the end of the fall semester, students in the sampled classes were asked to complete the New General Self-Efficacy (NGSE) scale (Chen, Gully, & Eden, 2001), and the CCR test-Form X (Ennis et al., 1964). The NGSE scale contains eight statements of general confidence in one's abilities, each measured on a scale from 1 (strongly disagree) to 5 (strongly agree). The NGSE inventory was selected over other instruments for (a) its desirable psychometric properties, including unidimensionality, construct validity, and reliability (Chen et al., 2001), (b) brevity, (c) clarity for the selected population (as assessed by pilot work), and (d) ability to capture students' underlying confidence to perform well across diverse tasks and situations, which is a motivational trait (Chen, Gully, Whiteman, & Kilcullen, 2000) that is positively related to other motivational traits, including need for achievement and conscientiousness (Chen et al., 2001). The CCR test-Form X is designed to measure 12 conditioning reasoning principles (see Table 1). The

Table 1
The Principles of Reasoning Tested by the CCR test-Form X

	Premise	Premise	Conclusion
1	If p, then q	P	Therefore q
2	If p, then q	Not p	Therefore not q
3	If p, then q	Q	Therefore p
4	If p, then q	Not q	Therefore not p
5	If p, then q	If q, then r	Therefore if p, then r
6	If p, then q		Therefore if not q, then not p
7	If p, then q		Therefore if q then p
8	p only if q	Not q	Therefore not p
9	p only if q	P	Therefore q
10	p if and only if q	Not p	Therefore not q
11	p only if q	Q	Therefore p
12	p only if q	Not p	Therefore not q

test contains 72 statements, 6 statements per principle. Each statement asks the reader to assume certain information (premise), and then decide whether a proposed statement is true (i.e., follows the premise), is false (i.e., contradicts the premise), or is indeterminable because there is not enough information to establish whether it is true or false (i.e., maybe). Instructions required participants not to guess, as well as to use only the information in each statement to select true, false, or maybe.

McLellan (2012) found that some questions of the CCR test-Form X referred to culturally unfamiliar content for United Arab Emirates (UAE) students. Pilot work with KSA students supported his findings. Thus, the content of statements 52, 66 (principle 5), and 62 (principle 10) was slightly modified to avoid unfamiliar terms and thus facilitate reading comprehension processes. Specifically, in statement 52, the unfamiliar terms (i.e., league pennant and hit a homer) of the baseball scenario were changed to those of a football scenario (i.e., prize and score a goal). In statement 66, the term "jumping rope" was changed to "running". Lastly, in statement 62, "marker" was used instead of "chalk". The alteration of linguistically and culturally opaque items had the desired effect of clarifying meaning since no questions arose regarding the modified items during administration. Although pilot work indicated adequate comprehension of test materials, students were instructed to seek clarification through the instructor or the translator function of their cell phones or laptops if an unfamiliar term was encountered. Questions rarely arose.

To ensure adequate time for in-class completion at the end of the semester, as well as minimize cognitive fatigue (a likely outcome of prolonged sustained attention), principles were randomly organized into four sets of three principles for a total of 18 statements (A, B, C, and D) per test-taker. McLellan (2012) reported that the average amount of time taken by UAE students to complete the whole test was 53 minutes. No measure of variability was reported. Our pilot work partially replicated McLellan's estimate with a range between 50 minutes and 70 minutes for whole-test completion. The option of breaking up testing time into separate periods of 20 to 30 minutes was considered as an alternative to segmentation of the test into smaller units. Because it was judged unfeasible by instructors, test segmentation was adopted to minimize disruption of ordinary class activities.

Each student completed a set. Random assignment was used to allocate sets to individual students. Each set was preceded by the practice questions included in the original test written by Ennis et al. (1964). Approximately a week later, students received feedback regarding their answers, and they were given the opportunity to discuss their choices with instructors.

Design

The study entailed a successive independent-samples design with condition as the between-subjects factor (baseline/formal instruction condition populated by freshmen versus post-instruction condition populated by sophomores). The key dependent measure was conditional reasoning performance. Self-efficacy scores, as well as GPA values, were factors whose potential contribution to performance was examined. A successive independent-samples design was chosen over a longitudinal design to avoid practice effects and to ensure timely delivery of performance feedback so that the test-taking activity could be treated as a learning exercise.

Results of Study 1

All results discussed in this section were considered significant if $p < .05$. Conditional reasoning scores were analyzed to answer each of the following questions:

Table 2
Mean Percentage Score of Each Principle of Reasoning Tested by the CCR test-Form X as a Function of Past Education Emphasizing Verbatim Learning (UAE and KSA) or De-emphasizing it (USA).

	UAE	**KSA**	**USA**	**KSA**	
	Whole Test	**Partial Test**	**Whole Test**	**Whole Test**	
Principles	English	English	English	English	Arabic
1	65.50	69.53	78.33	60.83	63.61
2	28.00	24.11	36.67	27.78	56.94
3	28.67	24.55	33.33	31.94	35.28
4	51.33	54.35	65.00	48.89	49.44
5	64.17	61.24	75.00	55.56	52.50
6	51.00	49.02	60.00	49.72	52.78
7	29.00	32.12	43.33	30.28	57.50
8	73.50	77.11	86.67	56.39	58.06
9	74.17	79.53	86.67	53.89	59.72
10	59.00	61.15	75.00	41.11	47.22
11	58.33	58.80	65.00	45.83	48.61
12	21.50	24.84	21.67	25.28	45.83
Mean	50.35	51.37	60.56	43.96	52.29
SEM	5.50	5.89	6.31	3.57	2.21

Note. Data of UAE students are from McLellan (2012), whereas those of USA students are from Ennis and Paulus (1965).

Does Knowledge of Conditional Reasoning Principles Differ Between Freshman and Sophomore KSA Students?

Overall performance (i.e., mean percentage correct scores collapsed across principles) of freshmen exposed to targeted critical thinking instruction were not significantly different from those of sophomores who, after exposure to such instruction, later attended an outcome assessment class where more generic and motivational critical thinking instruction was offered (M = 52.59% and M = 50.14%, respectively), F = 2.13, *ns*. The use of either GPA (M = 3.03, SEM = .06, and M = 3.10, SEM = .03), or self-efficacy (M = 3.76, SEM = .06, and M = 3.13, SEM = .05) as a covariate did not change this outcome, Fs ≤ 2.23, *ns*. Thus, evidence of stability of conditional reasoning knowledge (as measured by overall performance) from the freshman to the sophomore years, rather than loss or gain, was obtained.

Is Knowledge of Conditional Reasoning Principles Possessed by KSA Students Different from That Possessed by Other Students?

Performance pertaining to the 12 conditional reasoning principles was examined in an item analysis with sample as the factor and performance as the dependent variable. Samples, which included our students, UAE students (as reported by McLellan, 2012), and USA students (as collected by Ennis & Paulus, 1965), were intended to signify past educational experiences that emphasized verbatim learning (UAE and KSA) or deemphasized it (USA). It is important to note that performance (i.e., mean percentage correct on each principle) of the UAE sample included 361 Arabic-English bilingual college students majoring in business (ages 18-21), whereas the USA sample included 78 monolingual English-speaking high school students (age 17). These samples' data were used for comparison purposes as they constitute normative performance for the CCR Test-Form X. Table 2 reports descriptive statistics. Item analysis illustrated that at least one sample differed from another, $F(2, 22)$ = 43.00, MSE = 8.66, p <.001, η_p^2= .796. *LSD* pairwise comparisons indicated that the USA sample differed from the UAE and KSA samples, whereas the UAE and KSA samples did not differ from each other.

A complementary set of performance data involving mastery of principles was also utilized. Namely, the percentage of students who entirely (at least 5 or 6 items correct out of 6) or partially (at least 4 items correct out of 6) mastered each principle of reasoning. Table 3 reports descriptive statistics. This item analysis with sample as the factor and mastery as the dependent variable indicated that at least one sample differed from another, $F(2, 22)$ = 36.29, MSE = 23.16, p <.001, η_p^2= .767. Consistent with the earlier finding, *LSD* pairwise comparisons indicated that the USA sample differed from the UAE and KSA samples, whereas the UAE and KSA samples did not differ from each other. Thus, conditional reasoning scores of students whose earlier educational experiences

Table 3

Percentage of Students Who Partially or Entirely Mastered Each Principle of Reasoning Tested by the CCR test-Form X (at Least 4, 5 or 6 Items Correct out of 6) as a Function of Past Education Emphasizing Verbatim Learning (UAE and KSA) or Deemphasizing it (USA)

	UAE	**KSA**	**USA**	**KSA**	
	Whole Test	**Partial Test**	**Whole Test**	**Whole Test**	
Principles	English	English	English	English	Arabic
1	63	64	78	60	65
2	10	11	27	08	37
3	12	11	16	05	32
4	36	43	60	30	27
5	62	60	68	43	30
6	40	34	53	37	40
7	17	12	32	12	47
8	75	70	91	52	52
9	80	73	95	45	55
10	54	53	77	27	35
11	54	44	66	25	37
12	8	6	5	12	32
Mean	43	40	56	30	41

Note. Data of UAE students are from McLellan (2012), whereas those of USA students are from Ennis and Paulus (1965).

emphasized verbatim learning (UAE and KSA) were lower than those of students whose earlier educational experiences deemphasized verbatim learning (USA). The equivalent performance of UAE and KSA students could be interpreted as illustrating the negligible impact of cognitive fatigue caused by reading-comprehension processes engaged by a long test in a second language. The role of cognitive fatigue was further investigated in Study 2 in which we asked whether the same outcome could be reported in KSA students given the entire test.

What Does Contribute to Conditional Reasoning Performance of KSA Students?

A linear regression analysis was conducted to determine the relative contribution of condition, self-efficacy, and GPA to overall conditional reasoning performance. The only significant contribution to performance was made by self-efficacy (see Table 4). Since confidence in one's abilities translates into effort and persistence, it is not surprising that demanding cognitive operations, such as those of a conditional reasoning test, rely not only on test-takers' knowledge of key principles, but also on their self-efficacy.

Study 2

The results of Study 1 left open the possibility that students' cognitive overload, due to their using English in a challenging task, might depress performance. Reliance on a second language has been shown to negatively affect performance in other challenging tasks such as mathematical problems (Campbell et al., 2007) or calculations (Takano & Noda, 1993). According to Paas et al. (2003), working memory is limited in the amount of information that it can process. Thus, could students' overloaded working memory have prevented them from adequately processing the critical information of the items of the conditional reasoning test? Would taking the whole test magnify the hypothesized students' processing overload?

Method of Study 2

To investigate potential test language and length effects, 120 students in the post-instruction condition were given the whole test to complete within a timeframe of 2 hours with breaks initiated by students. Mean completion time was 1 hour. We selected 6 sections of learning outcome assessment out of 27 through convenience sampling (22% of the available classes). Random assignment determined for each student the language in which the test was written (60 students for language). Thus, Study 2 involved a cross-sectional design with language (Arabic and English) as the factor. Upon completion, participants were given feedback regarding their performance. Their self-efficacy score (Chen et al., 2001) gathered earlier was $M = 3.17$ ($SEM = .11$). Students' mean GPA was 3.09 ($SEM = .04$).

Three independent translators familiar with critical thinking constructs and instruments were recruited to

Table 4
Results of Regression Analysis of Conditional Reasoning Scores

Factors	*B*	*SE B*	ß	*t*
Study 1				
Condition	.40	1.66	.01	< 1
Self-Efficacy	3.23	.74	.21	4.37 *
GPA	1.08	1.21	.04	<1
Study 2				
Test Language	7.09	1.97	.31	3.59 *
Self-Efficacy	2.66	.86	.26	3.09 *
GPA	-0.39	2.04	-.016	< 1

Note. Study 1: R^2 = .045. Study 2: R^2 = .189. * Significant contribution to conditional reasoning performance.

ensure a culturally appropriate, native, and accurate Arabic translation. Dynamic equivalence, whose goal is naturalness of expression (Nida, 2004), was achieved through a consensus model (Scholz, Gutiérrez Doña, Sud, & Schwarzer, 2002), including back-translations, group discussions, and feedback from monolingual individuals (Sperber, 2004).

Results of Study 2

Is Performance Regarding Conditional Reasoning Principles Sensitive to the Language or Length of the Test?

To assess whether there was an effect of length or language of the test on the scores linked to the 12 conditional reasoning principles, an item analysis was conducted on the mean percentage correct scores of the 12 principles with sample as the factor. The samples considered were KSA students who had been given segments of the test (Study 1), as well as KSA students (Study 2) and USA students (Ennis & Paulus, 1965) who had completed the whole test (see Table 2). Following a significant effect of sample, $F(3, 33) = 6.24$, $MSE = 88.68$, $p = .002$, $\eta_p^2 = .362$, *LSD* pairwise comparisons indicated that when the language of the test was English, KSA students who took the whole test performed less well than those who took separate segments. KSA students taking the whole test in English also performed less well than USA students. Interestingly, KSA students who took the Arabic translation of the whole test performed better than KSA students who took the whole test in English, but less well than USA students.

An item analysis with sample as a factor was also conducted on scores involving mastery of principles to assess whether there was an effect of the test's language or length on the percentage of students who partially or entirely mastered the 12 conditional reasoning principles (see Table 3). Following a significant effect of sample, $F(3, 33) = 9.11$, $MSE = 150.76$, $p < .001$, $\eta_p^2 = .453$, *LSD* pairwise comparisons replicated with one exception in the patterns uncovered in the item analysis of mean percentage correct scores. The difference in mastery between KSA students taking the test in their first language (i.e., Arabic) and USA students was no longer significant, albeit in the expected direction.

These results indicated that completing a demanding test in one's first language or a shorter version of it focused on a few principles could considerably aid performance. Yet, language and length adjustments did not appear to be able to entirely compensate for past educational experiences (as illustrated by the higher performance of USA students).

What Does Contribute to Conditional Reasoning Performance of KSA Students Taking the Whole Test?

A linear regression analysis was conducted to determine the relative contribution of language, self-efficacy, and GPA to overall performance (mean percentage correct scores collapsed across all principles) of KSA students taking the whole test. The analysis indicated that the only significant contribution to performance was made by self-efficacy and language (see Table 4). In agreement with this analysis, taking the test in the second language was found to significantly lower performance compared with taking the same test in the first language, $F(1, 118) = 16.38$, $MSE = 120.23$,

p <.001, η_p^2= .122 (see last two columns of Table 2 for descriptive statistics). As per Study 1, GPA and self-efficacy treated as covariates did not eliminate this performance difference.

Discussion

The results of our investigation can be summarized in three key points: First, if test performance was measured as a difference between freshmen and sophomores, conditional reasoning knowledge remained largely stable as students' academic experience increased. Second, test-takers' self-efficacy beliefs and processing load (as determined by the need to sustain attention to the contents of a long test written in a second language) contributed to performance. Third, differences between USA and KSA students appeared to result from not only first versus second language processing and test length, but also past educational practices shaping the approach that students expressed towards the contents of the test.

The undergraduate curriculum to which our participants have been exposed promotes applications of critical thinking and highlights its utility. If the curriculum, as a whole, nurtures knowledge of critical thinking principles and shapes information processing accordingly, past educational experiences reinforcing verbatim learning may decrease in relevance and, to a certain degree, fade into disuse. In fact, students are expected to develop habits that are adaptive ways of coping with the demands of their lives. However, in KSA and UAE, college students tend to openly express a firm disposition towards verbatim learning as either a preference for veridical replication of information or aversion for alternative modes of information processing. Observations made by instructors at our university and others (Fareh, 2010; McLellan, 2012) support the idea of verbatim learning, both as a preference and as a habit that coexists with critical thinking. A disposition towards the former has been shaped and reinforced by a type of pedagogy widespread in schools of the Arab world which emphasizes the practice of memorization and recitation. Namely, students are expected to commit large portions of text to memory and are praised when they are able to reproduce encoded materials precisely (Iqbal & Ahmad, 2015). How can two contradictory approaches, one favoring verbatim learning and the other promoting active learning, coexist in our students? There are three aspects of habits that need to be considered if coexistence is to be understood: (a) the association between habits and preferences, (b) the utility of habits, and (c) the relationship between habits and norms (Lindbladh & Lyttkens, 2002). Verbatim learning is familiar and thus tends to be preferred over modes of active learning (as indicated by debriefing and pilot work data). In addition, in a few classes, such as Islamic Studies (I-IV) and oral communication (all courses of the general education curriculum at the university), memorization and recitation are considered key to performance, whereas in most other classes, norms, in the form of instructional requirements, render verbatim learning much less valuable. Thus, the habit of verbatim learning is not entirely discouraged in college and continues to exist in a few separate pockets of the academic curriculum.

The lower performance of UAE and KSA students relative to that of USA students highlights the relative importance of different cognitive factors. Namely, it brings to the forefront not only dispositions towards knowledge acquisition (as driven by students' educational past), but also sustained cognitive load (as illustrated by the need to maintain attention to the materials of a long test written in a second language). Pilot work alerted us to the relevance of test length. Although the original test instructions described the whole test as taking approximately 40 minutes, we found that students taking the test in their second language needed more time. For instance, in pilot work, mean completion time for the whole test was approximately 1 hour (without counting breaks). Students required pauses which lengthened individual test sessions. In McLellan's study (2012), mean completion time was 53 minutes. Significant performance differences of KSA students taking the whole versus the partial test in English were found in the analyses of Study 2. Clearly, segmental administration is less time consuming and burdensome to students. Thus, if group performance is of interest and available class time is limited, suitable administration may consist of portions of the test or the whole test translated in the first language of the test-takers.

Although the test items that McLellan described as difficult for UAE students and USA students (e.g., invalid statements) were also difficult for KSA students, performance was overall lower for UAE and KSA students, both of whom read test items written in English, their second language. In the Arabic version of the test, performance was more uniform. Second language processing is known to be cognitively demanding (Perani, & Abutalebi, 2005). Thus, to a certain extent, the lower performance of UAE and KSA students may be attributed to second language processing adding to the burden of a computationally demanding test (see Barrouillet & Lecas, 1999). Important to note here is that in both our research and that of McLellan (2012), students' English proficiency had been verified through standardized tests. Furthermore, at the time of testing students could clarify the meaning of unfamiliar terms through instantaneous translations. Of course, the fact that comprehension of statements can be quickly resolved through translation may not matter much if the act of

translating is distracting, further overloading available resources, and thus contributing to extra processing and fatigue. In support of the notion that second language processing may depress performance, McLellan (2012) found that UAE students perform more poorly than USA students, but similarly to students from the West Indies of similar age (Nolan & Brandon, 1984). Again, lower performance was associated with students whose first language was different from English (Arabic for UAE and Jamaican Creole for students from the West Indies; see Craig, 1980). Yet, one may argue that the comparison with USA normative data collected some time ago by Ennis and Paulus (1965) involves not only students with a different first language from that of KSA and UAE students, but also students who belong to a different generation. At the present time, the factors that may contribute to generational differences (e.g., basic knowledge, motivation, etc.) remain the realm of speculation. If such factors can be identified in future research, their impact may be either additive or multiplicative to the effects of linguistic factors.

The present study has several limitations that we hope to address in future research. For instance, the successive-independent design was selected over a longitudinal design to ensure prompt delivery of performance feedback to students and to satisfy instructors' demands for minimal disruption of class activities. The selected design did not allow us to assess the pre-intervention level of individual students' knowledge. However, a modest improvement of test performance from pre- (midsemester) to post-intervention (end of semester) was reported by Al-Ghamdi and Deraney (2013) in a longitudinal study conducted on the same population of students exposed to the same critical thinking curriculum and instruction but given a generic critical thinking test. The authors decided not to administer the pre-test at the start of the semester because freshmen's formal knowledge of critical thinking principles and terminology had yet to develop. By extrapolation, we can assume that our baseline students, whose knowledge was assessed at the end the semester, had acquired some formal understanding of conditional reasoning principles from the same critical thinking course and that such knowledge was preserved approximately a year later. Our pilot work supports this conclusion by showing weak formal knowledge of conditional reasoning principles in freshmen prior to their taking the critical thinking course of the general education curriculum. Because low test performance might lead to discouragement, frustration, and anxiety, all of which are detrimental to students' motivation in the critical thinking course in which they are about to enroll, the pre-testing of baseline freshmen was not entertained on a wide scale. Another limitation of the study is that only female students participated due to gender-segregation rules that prevented access to a male sample. Although McLellan (2012) found no evidence of gender differences in UAE students, Ennis and Paulus (1965) reported a minor difference favoring females in USA students. Thus, a further exploration of this issue may be warranted with samples of KSA students. Lastly, it is to be determined whether the experience of taking a conditional reasoning test, facing challenges, and receiving helpful feedback, if adequately conveyed and reinforced through class assignments and tests across the curriculum, may propel substantial changes in students' overreliance on verbatim learning. Successful habit formation in this area requires nurturing of alternative modes of thinking that are effortful and unfamiliar, but ultimately useful to students seeking to develop competence in the field of their choosing.

References

Al-Ghamdi, A. K. H., & Deraney, P. M. (2013). Effects of teaching critical thinking to Saudi Female university students using a stand-alone course. *International EducationStudies, 6*(7)*,* 176-188. doi:10.5539/ies.v6n7p176

Al-Wehaibi, H. U. (2012). Novel program to promote critical thinking among higher education students: Empirical study from Saudi Arabia. *Asian Social Science, 8*(11)*,* 193-204. doi:10.5539/ass.v8n11p193

Ausubel, D. P. (2012). *The acquisition and retention of knowledge: A cognitive view*. Berlin, DE: Springer Science & Business Media.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavior change. *Psychological Review, 84,* 191-215. doi:10.1037/0033-295X.84.2.191

Bandura, A. (1989). Human agency in social cognitive theory. *American Psychologist, 44*(9)*,* 1175-1184. doi:10.1037/0003-066X.44.9.1175

Bandura, A., & Schunk, D. H. (1981). Cultivating competence, self-efficacy, and intrinsic interest through proximal self-motivation. *Journal of Personality and Social Psychology, 41*(3)*,* 586-598. doi:10.1037/0022-3514.41.3.586

Barrouillet, P., & Lecas, J. F. (1999). Mental models in conditional reasoning and working memory. *Thinking & Reasoning*, *5*(4)*,* 289-302. doi:10.1080/135467899393940

Campbell, A. E., Adams, V. M., & Davis, G. E. (2007). Cognitive demands and second language learners: A framework for analyzing mathematics instructional contexts. *Mathematical Thinking and Learning, 9*, 3-30. doi:10.1080/10986060709336603

Campbell, T., Dollaghan, C., Needleman, H., & Janosky, J. (1997). Reducing bias in language assessment: Processing-dependent measures. *Journal of Speech,*

Language, and Hearing Research, 40(3), 519-525. doi:10.1044/jslhr.4003.519

Chen, G., Gully, S. M., & Eden, D. (2001). Validation of a new general self-efficacy scale. *Organizational Research Methods, 4*(1), 62-83. doi:10.1177/109442810141004

Chen, G., Gully, S. M., Whiteman, J. A., & Kilcullen, B. N. (2000). Examination of relationships among trait-like individual differences, state-like individual differences, and learning performance. *Journal of Applied Psychology, 85*(6), 835-847. doi:10.1037/0021-9010.85.6.835

Craig, D. (1980). Language, society and education in the West Indies. *Caribbean Journal of Education*, *7*, 1-17.

Cummins, D. D., Lubart, T., Alksnis, O., & Rist, R. (1991). Conditional reasoning and causation. *Memory & Cognition, 19*(3), 274-282. doi:10.3758/BF03211151

Douglass, S. L., & Shaikh, M. A. (2004). Defining Islamic education: Differentiation and applications. *Current Issues in Comparative Education, 7*(1), 5-18.

Eden, D. (1984). Self-fulfilling prophesy as a management tool. *Academy of Management Review, 9*(1), 64-71. doi:10.5465/amr.1984.4277938

Eden, D. (1988). Pygmalion, goal-setting, and expectancy. Compatible ways to boost productivity. *Academy of Management Review, 13*(4), 639-652. doi:10.5465/amr.1988.4307530

Ennis, R. H. (1964). A definition of critical thinking. *The Reading Teacher, 17*(8), 599-612.

Ennis, R. H. (1987). A taxonomy of critical thinking dispositions and abilities. In J. Baron, & R. Sternberg (Eds.), *Teaching thinking skills: Theory and practice* (pp. 9−26). New York, NY: Freeman.

Ennis, R. H. (2011). *The nature of critical thinking: An outline of critical thinking dispositions and abilities.* Paper presented at the 6th International Conference on Thinking of MIT, Cambridge, MA.

Ennis, R. H., Gardiner, W. L., Guzzetta, J., Morrow, R., Paulus, D., & Ringel, L. (1964). *The Cornell Conditional Reasoning Test, Form X*. Ithaca, NY: Cornell University.

Ennis, R. H., & Paulus, D. (1965). *Critical thinking readiness in grades 1-12: Phase I: Deductive reasoning in adolescence* (Report No: CRP-1680). Ithaca, NY: Cornell University.

Fareh, S. (2010). Challenges of teaching English in the Arab world: Why can't EFL programs deliver as expected? *Procedia-Social and Behavioral Sciences, 2*(2), 3600-3604. doi:10.1016/j.sbspro.2010.03.559

Halpern, D. F. (1998). Teaching critical thinking for transfer across domains: Disposition, skills, structure training, and metacognitive monitoring. *American Psychologist*, 53(4), 449-455. doi:10.1037/0003-066X.53.4.449

Hambleton, R. K., Merenda, P. F., & Spielberger, C. D. (2004). Issues, designs, and technical guidelines for adapting tests into multiple languages and cultures. In *Adapting educational and psychological tests for cross-cultural assessment* (pp. 15-50). New York, NY: Psychology Press.

Iqbal, J., & Ahmad, A. (2015). Effect of extensive rote learning experience on subsequent academic achievement. *Pakistan Armed Forces Medical Journal, 65*(4), 510-514.

Johnson, M., & Posner, G. J. (1971). *Testing the effect of verbal-quantitative aptitude discrepancy on the learning of deductive reasoning through programmed instruction. Final report* (Report No: 0EG-2-9-420054-1033). Albany, NY: The Research Foundation of State University of New York

Lehman, D. R., & Nisbett, R. E. (1990). A longitudinal study of the effects of undergraduate training on reasoning. *Developmental Psychology*, *26*(6), 952-960. doi:10.1037/0012-1649.26.6.952

Lindbladh, E., & Lyttkens, C. H. (2002). Habit versus choice: the process of decision-making in health-related behaviour. *Social Science & Medicine*, *55*(3), 451-465. doi:10.1016/S0277-9536(01)00180-0

Lun, V. M. C., Fischer, R., & Ward, C. (2010). Exploring cultural differences in critical thinking:Is it about my thinking style or the language I speak? *Learning and Individual Differences, 20*(6), 604-616. doi:10.1016/j.lindif.2010.07.001

McLellan, J. (2012). Assessing the critical thinking abilities of United Arab Emirates University business students. *Learning and Teaching in Higher Education: Gulf Perspectives, 6*(2), 1-17. Retrieved from https://ssrn.com/abstract=1932038.

Nickerson, R. S., Perkins, D., & Smith, E. E. (1985). *Teaching thinking*. Hillsdale, NJ: Erlbaum Associates.

Nida, E. (2004). Principles of correspondence. In L. Venuti (Ed.), *The translation studies reader* (pp. 126-140). New York, NY: Routledge.

Nisbett, R. E. (2013). *Rules for reasoning*. New York: NY: Psychology Press.

Nisbett, R. E, Fong, G. X, Lehman, D. R., & Cheng, P. W (1987). Teaching reasoning. *Science, 238,* 625-631. doi:10.1126/science.3672116

Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review, 108*(2), 291-310. doi:10.1037/0033-295X.108.2.291

Nolan, C. A., & Brandon, E. P. (1984, August). *Conditional reasoning in Jamaica*. Paper presented at the Harvard International Conference on Thinking, Cambridge, MA.

Oettingen, G. (1995). Cross-cultural perspectives on self-efficacy. In A. Bandura (Ed.), *Self-efficacy in changing societies* (pp. 149-176). New York, NY: Cambridge University Press.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist, 38*, 63-71.

Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research*, *66*(4), 543-578. doi:10.3102/00346543066004543

Pajares, F. (1997). Current directions in self-efficacy research. *Advances in Motivation and Achievement*, *10*(149), 1-49.

Pascarella, E. T., & Terenzini, P. T. (2005). *How college affects students* (Vol. 2). San Francisco, CA: Jossey-Bass.

Paul, R., & Elder, L. (2014). Critical thinking: Intellectual standards essential to reasoning well within every domain of human thought, Part 4. *Journal of Developmental Education*, *37*(3), 34-35.

Perani, D., & Abutalebi, J. (2005). The neural basis of first and second language processing. *Current Opinion in Neurobiology*, *15*(2), 202-206. doi:10.1016/j.conb.2005.03.007

Pithers, R. T., & Soden, R. (2000). Critical thinking in education: A review. *Educational Research, 42*(3), 237−249

Rugh, W. A. (2002). Arab education: Tradition, growth and reform. *The Middle East Journal*, *56*(3), 396-414.

Scholz, U., Gutiérrez Doña, B. G., Sud, S., & Schwarzer, R. (2002). Is general self-efficacy a universal construct? Psychometric findings from 25 countries. *European Journal of Psychological Assessment, 18*(3), 242-251. doi:10.1027//1015-5759.18.3.242

Sperber, A. D. (2004). Translation and validation of study instruments for cross-cultural research. *Gastroenterology, 126,* S124-S128. doi:10.1053/j.gastro.2003.10.016

Stipek, D. (1991). Characterizing early childhood education programs. *New Directions for Child and Adolescent Development*, *1991*(53), 47-55. doi:10.1002/cd.23219915308

Takano, Y., & Noda, A. (1993). A temporary decline of thinking ability during foreign language processing. *Journal of Cross-Cultural Psychology, 24*(4), 445-462. doi:10.1177/0022022193244005

Tiruneh, D. T., Verburgh, A., & Elen, J. (2014). Effectiveness of critical thinking instruction in higher education: A systematic review of intervention studies. *Higher Education Studies, 4*(1), 1-17. doi:10.5539/hes.v4n1p1

Weimer, M. (2002). *Learner-centered teaching: Five key changes to practice*. Hoboken, NJ: John Wiley & Sons.

SIDDIQUA AAMIR is a clinical psychologist and educator whose primary focus of research and practice is the education of women in changing societies. Her current research focuses on how language and culture shape learning. She received her Ph.D. in clinical psychology at the University of Peshawar (Pakistan).

RUNNA AL GHAZO is a special education researcher who has obtained her Ph.D. in rehabilitation from Southern Illinois University-Carbondale (USA). In addition to special education, her areas of interest include educational technology, curriculum and instruction, leadership, and technical writing pedagogy.

MAURA A. E. PILOTTI is a cognitive psychologist whose research interests include learning and memory processes across the lifespan. Currently, her research focuses on the interrelations of memory, language, and emotion. She received her Ph.D. in cognitive psychology at the City University of New York (USA).

HALAH ALKUHAYLI is a researcher of Islamic culture. Her expertise combines education and cultural studies. She earned her M.A. in Islamic culture from Imam Muhammad Ibn Saud Islamic University (KSA).