Critical Thinking Performance Increases in Psychology Undergraduates Measured Using a Workplace Recognized Test

Mercè Prat-Sala & Mike van Duuren

Department of Psychology

The University of Winchester

Corresponding author:

Merce Prat-Sala

Department of Psychology

The University of Winchester

Email: merce.prat-sala@winchester.ac.uk

Abstract

Background: Higher Education Institutions and Universities aim to provide students with a range of transferable skills that enable students to become more thoughtful and effective employees, citizen, and consumers. One of these skills is critical thinking.

Objective: The aim of the present research was to examine whether taking a psychology degree is concomitant with students' increase in critical thinking skills when students are not explicitly taught critical thinking.

Method: Study 1 utilized a cross-sectional design and Study 2 a longitudinal design. The Watson and Glaser Critical Thinking Appraisal (WGCTA^{uk}), was used to measure critical thinking.

Results: For both studies the overall scores of WGCTA, as well as scores of the subtest of *Recognition of Assumptions,* were significantly higher for final year than for first year students.

Conclusion: From the findings we conclude that the levels of critical thinking by final year psychology students may be enhanced.

Teaching Implications: We propose that teaching to other aspects of critical thinking such as *Evaluation of Arguments* and *Interpretation,* as measured by this test, could be beneficial in further developing psychology students' overall critical thinking performance.

*Keywords***:** critical thinking, transferrable skills, psychology undergraduates, WGCTA

Introduction

There are many skills that are required for the world of work including, information management, oral and written communication, collaboration, critical and analytical thinking, self-regulation, integrity and adaptability (e.g., Naufel et al., 2018). Given the importance of these skills not only in the workplace, but also to support almost any significant learning opportunity in life generally, many higher education institutions list these skills as part of the learning outcomes for their students (Rooh et al., 2019).

There is an argument that among these aforementioned skills-sets critical thinking is central for shaping effective global workforce roles (Liu et al., 2014). However, in addition to that in a world where the opportunity for the consumption of information and miss-information is ubiquitous, critical thinking becomes a valued skill for students to develop in order to participate as responsible members of their communities (Dam & Volman, 2004; Paul & Elder, 2019). Moreover, it is also a valued skill for differentiating between real and so called "fake news" (Musgrove et al., 2018; Paul & Elder, 2019).

Although interest in students' development of critical thinking has been longstanding, it is only relatively recent that empirical research has been focused on what factors might increase critical thinking skills. The majority of this work has focused on examining the effectiveness of critical thinking interventions (e.g., Barnet & Francis, 2012; Bensley et al., 2010; Cloete, 2018; Solon, 2007). However, very few empirical studies have focused on whether the learning, research and practice that students normally engage in, leads students to improve these skills (e.g., Liu et al., 2016a; Pascarella & Terenzini, 2005).

From a teaching perspective, two issues are relevant with respect to the teaching and assessing of critical thinking: (1) Teaching: whether critical thinking skills should be taught in the context of discipline specific matter and focus on cognitive skills relevant to that discipline; or whether universities should teach critical thinking skills with general vignettes

with non-discipline specific problem arguments and statements; and (2) Assessment: whether the tests used to assess students' critical thinking should be discipline specific or whether a generic critical thinking test should be used. The ability of students to transfer critical thinking skills to the workplace is important because if critical thinking skills are in the main more subject-specific, then the transferability of critical thinking skills to graduate employment might be more limited. Furthermore, given that many employers use generic tests (Watson & Glaser, 2018) perhaps this is an opportunity for Universities to demonstrate that the rhetoric that promises improved employability and transferable skills has actual relevance in relation to some of the criteria (occupational testing) used in the actual world of work.

Critical thinking is widely assumed to be an important part of psychology graduate, as reflected in the British Psychological Association (BPS) *Standards for the accreditation of undergraduate, conversion and integrated masters programmes in psychology* in the UK, and the American Psychological Association (APA) *Guidelines for the Undergraduate Psychology Major* (2013) in the US. Furthermore, psychology degrees include courses on research methods which promote rules and values of science such as objectivity, valid evidence, falsifiability and operationism; the use of a wide range of methods and statistical analyses; appropriate conclusions derived from empirical evidence and analysis (Bensley, 2009; Yanchar et al., 2008).

In this paper we present research which examines the extent, if at all, students taking a psychology degree, without any additional intervention or explicit teaching of critical thinking skills, further develop their critical thinking. In what follows, we review critical thinking literature that focuses on two aspects: the time span across assessment of students critical thinking skills; and the specific tests used to assess critical thinking skills. We then present two studies.

Empirical Evidence of Students' Development of Critical Thinking in Education

Time Span Between Testing

The focus of a considerable number of studies has been on the effectiveness of interventions to develop students' critical thinking skills in higher education, using a "pre- to post-test" design where the same participants are tested once at the beginning and again at the end of a semester (e.g., Bensley et al., 2010, Burke et al., 2014; Haw, 2011; Lawson, 1999; Lawson et al., 2015; Stark, 2012). In these studies, the time frame where critical thinking skills are expected to develop is relatively small ranging between 12 to 16 weeks. One of the main limitations of these studies is that they do not provide evidence on whether the gain in critical thinking, continues after post-testing at the end of a semester.

In contrast, the focus of other studies has been on possible student critical thinking gains over larger time spans than a single semester (e.g., Bauwens & Gerhard, 1987; Behrens, 1996; Cloete, 2018; Liu et al., 2016a, 2016b; Roohr et al., 2019) using either cross-sectional (i.e., data from two different cohorts of students, e.g., first year vs. final year students) or longitudinal designs (i.e., data from the same group over time).

Using a longitudinal design, the authors of several studies on nursing training programs have found a lack of gain in critical thinking skills (e.g., Bauwens & Gerhard, 1987; Behrens, 1996; Jones & Morris, 2007, see also Hubert & Kuncel's, 2016, meta-analysis). However, when including participants from a range of disciplines, Roohr et al. (2017) found a significant difference in students' critical thinking scores after four/five years in higher education, but not after one, two or three years, suggesting the need to measure critical thinking skills over a longer period of time.

Using a cross-sectional design and students from different American colleges and different disciplines, Liu et al.'s (2016b) findings suggest that scores from final year students

were significantly higher than the first years' scores. Similar findings were obtained by Mines et al. (1990) who include mathematics and social science students from a US University.

Subject-Specific vs. Generic Critical Thinking Tests

Over the past decades, many definitions of critical thinking have been put forward (e.g., Ennis, 1987; Facione, 1990; - see also Hitchcock 2018, Griggs et al., 1998 and Liu et al., 2014 for reviews of definitions). Following a panel of experts on critical thinking, Facione (1990, p. 3) offered the following broad definition of critical thinking with its emphasis on cognitive skills: "We understand critical thinking to be purposeful, self-regulatory judgment which results in interpretation, analysis, evaluation, and inference, as well as explanation of the evidential, conceptual, methodological, criteriological, or contextual considerations upon which that judgment is based."

One of the controversies in the critical thinking literature is whether the test used to measure critical thinking skills should be subject-specific or a generic critical thinking test. Some researchers have developed subject-specific critical thinking tests, that focusing on the content and cognitive skills related to a particular discipline (e.g., Bensley et al., 2010; Lawson, 1999; Lawson et al., 2015; Wentworth & Whitmarsh, 2017). For example, Lawson et al. (2015) developed further the Psychological Critical Thinking Exam (PCTE) proposed by Lawson (1999) which they claim taps into the critical thinking cognitive skills of "*Evaluation of claims*" (see also Wentworth & Whithmarsh, 2017, for another psychology specific test with similar characteristics to the PCTE). However, one of the main limitations of these studies is that it is not clear the extent to which the skills developed for subject-specific discipline tap into non-discipline specific critical thinking.

Alternatively, some studies have focused on students' level of critical thinking using generic critical thinking tests such as WGCTA or Cornell Critical Thinking Test (CCTT). Tests such as WGCTA are called generic because they assess the individual's critical thinking

skills applied to statements that reflect the wide variety of arguments from many everyday life situations such as newspapers, magazines, conversations and media material in general (Watson & Glaser, 2018).

The findings of the studies using WGCTA or CCTT are mixed. Some show that following an intervention, students' performance in critical thinking skills increase (e.g., Barnett & Francis, 2012; Cloete, 2018; Solon, 2007) while other studies shows no gain between pre- and post-test (Renaud & Murray, 2008). Furthermore, Cloete (2018) showed that both the control and the experimental group improve their performance on the WGCTA test, albeit that the improvement in the control group was more limited, thereby showing that students were able to improve their critical thinking performance without intervention. Burke et al. (2014) also used the WGCTA test to assess students' critical thinking skills. The aim of their study was to compare directly psychology and philosophy undergraduates' performance (cf. Ortiz, 2007). Findings revealed that only the philosophy students improved their critical thinking between pre- and post-testing. These authors argue that the lack of improvement in critical thinking skills for the psychology students might have been due to psychology students learning mainly inductive reasoning (statistical and methodological), while WGCTA is mainly (but not exclusively) based on deductive reasoning. Nevertheless, what is unclear from their study is whether there is any long term gain in critical thinking performance as a result of their degree experience, as participants were tested after (only) 15 weeks.

Aim of Study 1

The aim of study 1 was to examine whether students taking an accredited psychology degree[1] without additional explicit teaching of critical thinking skills (e.g., Bensley et al., 2010; Haw, 2011; McLean & Miller, 2010; Stark, 2012) gain in critical thinking skills. We

[1] In the UK an accredited psychology degree includes the curriculum specified by the British Psychological Association (BPS), which states that students must undertake several courses on research methods, quantitative and qualitative analyses, as well as courses on the core area of psychology, such as development psychology, cognitive psychology, social psychology, historical and conceptual psychology and individual differences.

used the critical thinking test WGCTA[uk] principally because (1) it is a recognized psychometric test with psychometric properties (Watson & Glaser, 2018, 2002); (2) it assesses the individual's critical thinking skills applied to statements that reflect the wide variety of material encountered in many everyday life situations (Watson & Glaser, 2018).

Study 1 compared a group of first year psychology students at the beginning of their degree with a group of third year students at the end of their degree[2]. We had two research questions: (1) under the assumption that psychology courses enable students to develop their critical thinking, it was hypothesized that Year 3 students would show higher scores in the WGCTA critical thinking test than the Year 1 students, when at the same time controlling for potential confounds such as age, entry qualification and socio-economic background. (2) Burke's et al. (2014) argue that psychology degrees address mainly inductive critical thinking. Furthermore, as far as we know, to date, research has not included WGCTA subtest analysis. With the second research question we aimed to explore whether Year 3 performance scores on each of the subtest differed from Year 1 performance scores.

Study 1: A Cross-Sectional Study

Method

Participants

Participants were 184 psychology students from a University in England, UK. Ninety-four participants were first year students (15 male, 79 female) mean age 19.47 (range 18-44), and 90 were third year students (17 male, 73 female) mean age 21.30 (range 20-33). Deprivation index[3] (DEPI) based on the ZIP code of first year student's home indicated that

[2] In the UK the majority of degrees, such as psychology, are three years. Students in their first year (Year 1) are in their first year at University-the equivalent of United States freshman. Students in their third year (Year 3) are in their final year of their degree -the equivalent of United States senior.

[3] Higher Education Access Tracker (HEAT) is a nationally utilized widening participation profiler which analyses student zip data analysis using an algorithm which takes account of a wide range of information including that from UK education funding bodies, parental income, employment, health, education and skills, barriers to housing and services, living environment and housing. HEAT scale ranges from 0-4 with 4 indicating highest deprivation values and 0 the lowest.

23.9% came from deprived areas (scale 2-4) while 76.1% came from non-deprived areas (scale 0-1). For third year students the index indicated that 20.2% came from deprived areas (scale2-4) while 79.8% came from non-deprived areas (scale 0-1). We did not collect information on race/ethnicity of the participants however the student population at the University was 90-92% European-white.

The entry qualifications of participants included A-levels, BTec, HND, International Baccalaureate (IB), and Access routes, with a predominance of A-levels (see Table 1 see below). A-Level is a qualification offered by education institutions in the UK and which is used by Universities to assess a student's eligibility for an undergraduate degree course. International students are more likely to study an International Baccalaureate as a form of access to a UK University, and mature students take the Access route, which is a qualification that prepares individuals without a traditional qualification for study at University. BTec and HND are diplomas in further education and vocational qualifications[4]. We had full details of students entry points for 89.5% Year 1 and 96.% Year 3 students. An independent t-test between these two groups showed no significant differences on level points between Year 1 (M = 87.01, *SD* = 15.33) and Year 3 (M = 90.49, *SD* = 13.15); $t(169) = 1.57$, p = .118, CI [7.739 - .877, Cohen's d = 0.24, small effect size difference, Cohen, 1988).

INSERT TABLE 1 HERE

Measures

Students' critical thinking was measured using the Watson-Glaser Critical Thinking Appraisal (WGCTA^{UK}, 2002), a measure of verbal reasoning the content of which relates to a wide range of contemporary socio-political and everyday scenarios which are presented in the form of problems, statements, arguments and interpretations. This 80-item instrument

[4] BTEC Diploma is awarded after completion of a program that is more vocational than academic and is comparable to completion of a vocational senior high school program in the United States. HND is similar to Associate's degree in the US.

consists of five different subtests, each consisting of 16 items, measuring the following aspects of critical thinking: *inference, recognition of assumptions, deduction, interpretation,* and *evaluation of argument* (Watson & Glaser, 2002, 2018). See Appendix A for full definitions of each subtest and example items.

The test as a whole measures a candidate's critical ability to correctly identify answers which are correct in an absolute-logical way, as well as those for which only a more probabilistic judgment can be given, given the sufficiency or otherwise of evidence provided.

The WGCTA^{uk} raw scores range from 0-80. The data reported below are reported in terms of raw scores adjusted for chance[5] (see Wagner & Harvey, 2006) with higher scores indicating a greater facility for critical thinking, or if indicated, in terms of standardized scores which are required for any comparison between different sample data previously obtained (WGCTA^{uk} 2002).

Procedure

To ensure that testing took place in conditions commensurate to psychometric testing for job recruitment, the WGCTA^{UK} was administered strictly in accordance with the procedure outlined in the manual (Watson & Glaser, 2002, 2018) by qualified BPS occupational test administrators. One of the test administrators was responsible for the test score conversion.

On arrival participants were seated and were asked to read the briefing instructions and asked for written consent. The test instructions were then read out with an opportunity for asking questions. Participants were given a maximum of 50 minutes to complete the test. The test was administered in paper-and-pencil form.

[5] As the participants had a binary choice for answer to four of the scales (see Appendix A), all raw scores were adjusted for chance (see, Watson & Glaser, 2002; Wagner & Harvey, 2006).

First year students completed the WGCTA^{UK} during the first 5 weeks of their first semester at University, which meant that they had not completed any credits at the University, whereas third year students did so during the last 4 weeks of their final semester. These third year student had all completed 285 credits in psychology.[6] To encourage participation first year students were offered credit participation points towards their required psychology study participation quota. All final year students, in recognition of their time commitment, were offered a small financial incentive (£10.00) to take part.

Results

Concerning research question 1, it was predicted that overall WGCTA^{uk} scores would be higher for Year 3 than Year 1 students. We present de descriptive analyses in Table 2. To examine whether the performance scores from students in Year 3 were significantly higher than from Year 1 students, we carried out an Analysis of Covariance ANCOVA with overall WGCTA as the dependent variable, year group as the fixed factor and age, A-Level score and DEPI score as covariate. The main effect of year group was significant $F(1, 149) = 3.947$, $p = .049$, $\eta_p^2 = .026$, CI [52.943, .142]) indicating that students in Year 3 had significantly higher scores than students in Year 1.The covariate of age ($F(1, 149) = .371$, $p = .543$, $\eta_p^2 = .002$, CI [11.173, 5.906],) entry qualification ($F(1, 149) = 3.614$, $p = .059$, $\eta_p^2 = .024$, CI [.026, 1.367]) and levels of deprivation were not significant ($F(1, 149) = .092$, $p = .763$, $\eta_p^2 = .001$, CI [7.882, 5.788]). Cohen (1988) indicate that the effect size benchmark for a partial eta squared is defined as small = 0.01, medium = 0.06 or large = 0.14.

INSERT TABLE 2 HERE

With the research question 2, we aimed to explore whether the performance scores on the WGCTA subtests were different between Year 3 and Year 1 students. Table 2 shows the

[6]Students on an accredited single honours degree in psychology in the UK take courses in psychology only. This differs from the United States degrees based on the Liberal Arts philosophy where students take courses from a range of subjects to get a broad educational foundation before specialising on their majors.

descriptive statistics (means and standard deviations) of subscale scores for both groups. To analyze whether any of the differences were statistically significant further ANCOVAS were carried out, one for each of the WGCTA subscales, using the subscales scores as the dependent variable and year group (Year 1 vs. Year 3) as the fixed factor. Students' age, scores on entry qualification and DEPI scores were entered as covariate. As can be seen from Table 3, there was a significant difference between Year 1 and Year 3 on both the measures of *Inference* and *Recognition of Assumptions*. There was no significant difference between the year groups on the other three WGCTA subscales. The covariate age was not significant for any of the subscales. The covariate entry scores was significant for the subscale of *Recognition of Assumptions* only and the covariate DEPI was significant for the subscale *Interpretation* only.

INSERT TABLE 3

Study 1 Discussion

The findings of study 1 support the hypothesis that final year students evidence a higher level of critical thinking skills than students at the beginning of their degree course using an industry standard critical thinking test. These findings are comparable with other research which included students from other disciplines and utilized a cross-sectional design (e.g., Liu et al., 2016b; Mines et al., 1990) and builds on previous research that has also used the WGCTA test (e.g., Behrens, 1996; Burke et al., 2014; Cloete, 2018; McLean & Miller, 2010; Mines et al., 1990). These findings suggest that participating in a psychology degree at University could increase students' critical thinking skills. The findings relating to specific subscale scores also extend previous findings that have focused only on the *overall* WGCTA scores (e.g., Behrens, 1996; Burke et al., 2014; Cloete, 2018; McLean & Miller, 2010; Mines et al., 1990) to provide more details on the subscale scores for psychology students. In particular, it shows that only *Inference* and *Recognition of Assumptions* were significantly

different between the two groups, partly confirming the argument made by Burke et al. (2014).

However, there are some important limitations to the above study that lead us to treat the results with caution. For this study we followed a cross-sectional design comparing two group in two different stages of their degree. Although, here, we include age, entry qualification scores (A-Level scores) and deprivation information (DEPI scores) as covariates to take into account possible confounding variables affecting the main variables, the two groups of students were nevertheless not matched for other variables that could have affected the results, such as for example, motivation, intellectual skills, cognitive ability, academic performance or critical thinking disposition (cf. Bensley et al., 2010; Faccione, 1990; Solon, 2007).

This led us to carry out study 2 using a longitudinal design where we compare the scores of students in Year 1 with the scores of the *same* students at the end of their studies, in Year 3. The longitudinal design allowed for some potentially confounding variables (such as, for example, Entry Point and DEPI, as well as intellectual skills, cognitive ability and critical thinking disposition) to be controlled for by including the same participants.

The aims of this second study were twofold: (1) to examine whether we could replicate the findings of study 1 using a longitudinal design. Under the assumption that undergoing a psychology curriculum enables students to develop their critical thinking, it was hypothesized that the performance scores of the students in their final year would be higher than the scores they obtained at the beginning of their study. Furthermore, following the results of study 1, it was predicted that the scores of the subtests of *Recognition of Assumptions* and *Inference* would be higher in Year 3 than in Year 1; (2) in order to be able to provide some possible recommendation regarding teaching to further enhance students' performance in critical thinking we examined the contribution of each subtest of the

WGCTA to the final overall WGCTA scores. To this aim, we carried out a hierarchical regression analysis.

Study 2: A Longitudinal Perspective

Method

Participants

The same participants who were Year 1 students at the time of Study 1 took part in Study 2. From the initial 94 students who took part in study 1, 63 took the WGCTA^{uk} again when they were in Year 3 (10 males and 53 females).

Measures and Procedure

The measures and procedure were the same as for Study 1. As these students were now in Year 3 of their study, as alternative to course credits as those would not have been of benefit to them at the end of their degree, they received £10.00 for taking the time to take part in the study. The same qualified BPS occupational test administrators were employed for data collection and score transformation.

Results

The first aim of this second study was to examine whether the students' scores for Year 1 differed from the scores of the same students in Year 3. Table 3 shows the descriptive statistics (means and standard deviations) of scores obtained during Year 1 and Year 3. As can be seen the overall WGCTA^{uk} scores were higher for Year 3 than for Year 1.

INSERT TABLE 4 HERE

To assess whether these differences were statistically significant, we carried out a repeated measure ANCOVA with age at Year 3 as covariate and total WGCTA scores as repeated measures. From the results we found that there was a significant difference in the overall critical thinking performance between students in their first and final year $F(1,61) =$

6.029, $p = .017$, $\eta_p^2 = .090$. The covariate of age was not significant ($F(1, 61) = 3.821$, $p = .06$, $\eta_p^2 = .059$.

In addition, to examine whether the scores of each of the 5 subtests reliably differed between study years, further repeated measure ANCOVAs were carried out, with age at Year 3 as covariate and the scores on the WGCTA subtest as repeated measures. From the findings it transpired that there was a significant difference between Year 1 and Year 3 for *Recognition of Assumptions* $F(1,61) = 3.989$, $p = .050$, $\eta_p^2 = .061$. The covariate of age was not significant $F(1, 61) = 2.600$, $p = .112$, partial $\eta_p^2 = .041$.

There was no significant difference, however, for the remaining subtests of critical thinking included in the WGCTA: *Inference* $F(1,61) = 0.020$, $p = .888$, $\eta_p^2 = .001$; the covariate age $F(1,61) = 0.099$, $p = .754$, $\eta_p^2 = .002$; *Deduction* $F = (1,61) = 1.966$, $p = .1.66$, $\eta_p^2 = .031$; covariate age $F(1,61) = 0.099$, $p = 1.696$, $\eta_p^2 = .027$; *Interpretation* $F(1,61) = 0.634$, $p = .429$, $\eta_p^2 = .010$; covariate age $F(1,61) = 0.503$, $p = .481$, $\eta_p^2 = .008$; or *Evaluation of Arguments* $F(1,61) = 1.651$, $p = .204$, $\eta_p^2 = .021$; covariate age $F(1,61) = 1.265$, $p = .265$, $\eta_p^2 = .020$.

The second aim of the study 2 was to examine the data as a means to provide possible recommendations regarding teaching to further enhance students' performance in critical thinking. In particular we focused on how much each WGCTA subtest contributed to the overall WGCTA scores when students were in their final year of study. To this effect we carried out a hierarchical regression analysis with overall score as the dependent variable[7].

INSERT TABLE 5

[7] On the first step we included age of participants, entry qualification scores and deprivation information (DEPI scores) as fixed factors. In the second step we entered the 5 sub-scales of the WGCTA (Year 3) using a stepwise method.

When age, DEPI and entry qualification scores were included as independent variables in the first step, a non-significant model emerged $F(3,51) = 0.871$, $p = .468$, $R^2_{Adj} = .007$.

A significant model emerges when the subscale of *Deduction* was included as independent variable $F(4,50) = 27.204$, $p = .0001$, $R^2_{Adj} = .660$ accounting for 66% of the variance, with a further significant model emerging when *Inference* was included ($F(5,49) = 57.013$, $p = .0001$, $R^2_{Adj} = .838$) accounting for a further 17.8% of the variance. When *Recognition of Assumptions* was entered as independent variable, a significant model also emerged $F(6, 48) = 115.624$, $p = .0001$, $R^2_{Adj} = .927$, accounting for a further 8.9 % variance, as was the case when *Evaluation of Argument* was included $F(7,47) = 169.286$, $p = .0001$, $R^2_{Adj} = .956$, accounting for a further 2.9% of variance. Finally, when *Interpretation* was included a final significant model emerged $F(8,46) = 1049.889$, $p = .0001$, $R^2_{Adj} = .995$, accounting for a further 3.8% of variance.

Study 2 Discussion

The findings of study 2 are similar to those of study 1. Students in Year 3 reported higher scores on the overall WGCTA scores and hence overall higher levels of critical thinking skills than when they were at the beginning of their degree and had yet to complete any credits at University. The results are also similar to those found for study 1 with respect to the measures of *Recognition of Assumptions*.

Furthermore, the findings of Study 2 extend to psychology students previous results which utilized a longitudinal design with nursing degree students (e.g., Bauwens & Gerhard, 1987; Behrens, 1996; Jones & Morris, 2007) or students from arts, humanities and science disciplines (e.g., Roohr et al., 2017).

Finally, the results of the hierarchical regression analysis provide us with an overview of how much each individual subtest contributes towards the overall WGCTA score. Looking

at the literature, research on critical thinking which employs the WGCTA as a measure of critical thinking focuses on the overall scores without considering its component measures. Whereas overall scores are used by companies and organizations which utilize the WGCTA as a tool for selection or career advancement of employees (Watson & Glaser, 2002, 2018) from an education perspective, it is important to know what contributes to the overall WGCTA score in order to provide any recommendation to help further psychology students to improve their critical thinking. The findings of the hierarchical regression analysis suggest that the component measures which contribute least to the overall WGCTA are *Evaluation of Arguments* and *Interpretation*. This may suggest that there is value in addressing the skills underlying these specific measures more explicitly, from a teaching and instruction point of view (see below), in order to advance student critical thinking.

General Discussion

The aim of the present research was to investigate the extent to which taking a psychology degree is concomitant with students' increase in critical thinking skills where students are *not* explicitly taught critical thinking but develop critical thinking skills through the learning experience traditionally characteristic of psychology courses.

Findings of both the cross-sectional and longitudinal studies reported above were similar. With respect to the overall WGCTA scores, students in their final year of their degree obtained significantly higher WCGTA performance scores than students at the start of their degree. Overall, the findings of our studies are consistent with, and extend to psychology, the claims made by Pascarella and Terenzini (2005) who suggest that attending university on the whole improves students' critical thinking performance (see also Huber & Kuncel, 2016).

Our results also extend previous findings that have included psychology students as participants and a subject-specific critical thinking test (e.g., Lawson, 1999; Lawson et al.,

2015; Haw, 2011) to the WGCTA critical thinking test and by assessing participants increase in critical thinking skills for longer than one semester.

At the same time, our findings contrast with previous longitudinal research with students in nursing programs which found no increase in critical thinking skills (e.g. Behrens, 1996; Bauwer & Gerhard, 1987; Mines et al., 1990). They also contrast with previous research that investigated a performance increase from a specific intervention (e.g., Renaud & Murray, 2008; Stark, 2012; Williams, et al., 2004;).

With respect to the WGCTA subtests, from both studies, we found that students' critical thinking performance increases for *Recognition of Assumptions*. Very few studies have looked at WGCTA subtest differences and the few that exist have looked at the relationship between WGCTA subtest and grade scores (Gadzella et al., 2002; Steward & Al-Abdulla, 1989). For example, Gazella et al. (2002) found that only the subscales of *Interpretation* and *Evaluation of Arguments* significantly predicted GPA.

Implications of the Results

Graduates need to demonstrate a range of skills and attributes to successfully enter the workplace, with better paid jobs frequently going to candidates with better critical thinking measures. There are a very large number of occupations, including occupations attractive to psychology graduates, where critical thinking is included as a basic skill required for these occupations (e .g. Cottrell, 2017; O*Net Online, 2019, Liu et al., 2014) and there is evidence of a significant high positive correlation between the WGTCTA overall scores and job performance (Pearson-LatentLens, 2016; The Bar Standards Board, 2015). The higher levels of critical thinking skills psychology students' evidence at the end of their degree compared to the start of their course may highlight the relative value of taking a psychology degree in terms of employability.

However, the need for a thorough command of critical thinking is not confined to the world of employment, rather it relates to our ability to make sense of information flow generally, where it is important that we are able to distinguish between sound or cogent arguments and so called "fake news" information and other forms of misinformation (Musgrove et al., 2018; Paul & Elder, 2019; Pennycook & Rand, 2019). For example, Pennycook and Rand (2019) show that participants' scores on analytical reasoning were positively correlated with the ability to differentiate between "fake news" and real news. Furthermore, critical thinking is a valued skill for students in order to develop good citizenship (Dam & Volman, 2004; Paul & Elder, 2019).

From the results of the hierarchical regression analysis reported above, we can see that the WGCTA subtest of *Evaluation of Arguments* contributed a modest 2.9% towards the overall WGCTA score. *Evaluation of Arguments* requires the individual to analyze the evidence and arguments put forward in a text (or conversation). To this aim, it is important to be objective and work logically through the arguments and information put forward. There is the need to distinguish between arguments appealing to logic rather than emotion and avoid privileging information which confirms a preferred perspective. According to Lawson et al. (2015) the Psychological Critical Thinking Exam (PCTE) taps into the critical thinking skills of "*Evaluation of claims*". The work of Lawson et al. shows that over the span of a 15-week semester, senior psychology majors scored significantly higher on the PCTE than junior psychology majors, senior biology majors, senior art majors, and introductory psychology students. Other interventions focusing on the increase of argument development have also found positive results (e.g., Hasnunidah et al., 2015; Kuhn & Udell, 2003). Although the long-term impact of these interventions is unknown, teaching psychology undergraduates explicitly how to critically evaluate arguments might help increase their overall critical thinking skills.

Similarly, from the results of the hierarchical regression analysis in study 2 we can see that the WGCTA subtest of *Interpretation* contributed 3.8% towards the overall WGCTA score. According to Facione (1990) interpretation means to understand the meaning or significance of information. Students practice the skill of interpretation when they comprehend and express the meaning or significance of a wide variety of experiences (Facione, 1990). Teaching students to effectively interpret text would help students towards developing interpretation critical thinking skills. Furthermore, the use of collaborative or peer-learning has been reported to enhance critical thinking (Gokhale, 1995), including interpretation.

Limitations and Future Directions

One limitation of the studies presented here is that the higher performance levels of critical thinking at the end of the course could be due to students' maturation (see e.g., Huber & Kuncel, 2016; Pascarella & Terenzini, 2005) and not due to reading a degree in psychology. To statistically control for this, we used age as a covariate for the analyses for both study 1 and study 2 and included it in the first step of the hierarchical multiple regression analysis. From the results we concluded that age did not affect the significant difference between the overall critical thinking scores of Year 1 and Year 3 students. In this context Pascarella and Terenzini (2005) have claimed that when studies control for maturation, University attendance still produces significant gains.

Our studies did not include participants taking a degree other than psychology and hence our results cannot be generalized to students taking other disciplines. This raises the issue of the lack of a control group to compare psychology students' findings with. In other words, whether the findings in our studies are related to students taking psychology rather than some other relevant variable. However, this raises the question of what constitutes a relevant control group in this context. Including students from other disciplines could

potentially reveal the relative efficacy, or otherwise, of the experience of a psychology curriculum in this setting. However, one could argue that by including students from different disciplines to compare the relative efficacy of different curricula in this regard does not provide the best control measure either in that a better control group might be similarly aged individuals not pursuing a degree course. But then, as Huber and Kuncel (2016) argue, it is difficult to conduct such study comparing individuals attending or not attending University while at the same time controlling for many other potential confounds. Future research could include students enrolled in different disciplines to examine the relative efficacy of enhancing students' critical thinking in each discipline.

The sample size utilized here was comparable to many studies that have examined students' gains in critical thinking utilizing either a cross-sectional or longitudinal design (e.g., Jones & Morris, 2007; McLean & Miller, 2010; Mines et al., 1990; Roohr et al., 2017). Nevertheless, taking into consideration the large population of students in our higher education institutions, the sample is relatively small. Furthermore, issues of sampling are particularly relevant for the longitudinal study as participants who took the test twice were by and large a much more self-selected group that the students who partook in the study only once at the beginning of their degree. Another limitation is the small number of males in our sample, which could imply that current findings might be restricted to female psychology students only. Finally, the participants in our studies came from only one institution with the same aims and ethos with respect to students learning and development. Hence further research with a large number of participants and more institutions are needed to confirm present findings.

Conclusion

From the findings we conclude that the scores of students taking a psychology degree were significantly higher in Year 3 than Year 1 and hence there was an enhancement of their

critical thinking performance, even when critical thinking was not explicitly taught, as measured with WGCT, an industry standard psychometric test. We suggest that to further increase psychology students' critical thinking skills instructors might focus on the development of skills related to the WGCT component measures of *Evaluation of Arguments* and *Inference*.

References

American Psychological Association. (2013). *APA guidelines for the undergraduate psychology major: Version 2.0.* Washington, DC. Retrieved September, 15, 2019, from http://www.apa.org/ed/precollege/about/psymajor-guidelines.pdf

Barnett, J., & Francis, A. L. (2012). Using higher order thinking questions to foster critical thinking: A classroom study. *Educational Psychology, 32*(2), 201-211. http://doi.org/10.1080/01443410.2011.638619

Bauwens, E. E., & Gerhard, G. G. (1987). The use of the Watson-Glaser critical thinking appraisal to predict success in a Baccalaureate Nursing Program. *Journal of Nursing Education, 26*(7), 278-281. http://doi.org/10.3928/0148-4834-19870901-6

Behrens P. J. (1996). The Watson-Glaser critical thinking appraisal and academic performance of diploma school students. *Journal of Nursing Education,35*(1), 34-36. http://doi.org/10.3928/0148-4834-196601-10

Bensley, D. A. (2009). Thinking critically about critical thinking approaches: Comment on Yancher, Slife, and Warne (2008). *Review of General Psychology, 13*(3), 275–277. http://doi.org/10.1037/a0015654

Bensley, D. A.; Crowe, D. S., Bernhardt, P., Buckner, C., & Allman, A. L. (2010). Teaching and assessing critical thinking skills for argument analysis in Psychology. *Teaching of Psychology, 37*(2), 91-96. http://doi.org/10.1080/00986281003626656

Burke, B. L., Shears, S. R., Kraus, S., & Roberts-Cady, S. (2014). Critical analysis: A comparison of critical thinking changes in psychology and philosophy classes. *Teaching of Psychology, 41*(1), 28-36. http://doi.org/10.1177/0098628313514175

Cloete, M. (2018). The impact of an integrated assessment on the critical thinking skills of first-year university students. *Accounting Education, 27*(5), 479-494. http://doi.org/10.1080/09639284.2018.1501717

Cohen J. (1988). *Statistical power analysis for the behavioral sciences*. Routledge.

Cottrell, S. (2017). *Critical thinking skills: Effective analysis, argument and reflection*. Palgrave-Macmillan.

Dam, G. & Volman, M. (004). Critical thinking as a citizenship competence: teaching strategies. *Learning and Instruction, 14*, 359-379. https://doi.org/10.1016/j.learninstruc.2004.01.005

Dick, R. D. (1991). An empirical taxonomy of critical thinking. *Journal of Instructional Psychology*, *18*(2), 79-92.

Ennis, R. H. (1993). Critical thinking assessment. *Theory Into Practice, 32*(3), 179-186. http://doi.org/10.1080/00405849309543594

Facione, P. A. (1990). Critical Thinking: A statement of expert consensus for purposes of educational assessment and instruction. A Delphi Report. E*ducational Resources Information Center (ERIC).*

Gadzella, B. M. Baloglu, M., Stephens, R. (2002). Prediction of GPA with educational psychology grades and critical thinking scores. *Education, 122*(3), 618-623.

Gokhale, A. A. (1995). Collaborative learning enhances critical thinking. *Journal of Technology Education, 7*(1). Retrieved February 24, 2020, from https://scholar.lib.vt.edu/ejournals/JTE/v7n1/gokhale.jte-v7n1.html.

Haw, J. (2011). Improving psychological critical thinking in Australian university students. *Australian Journal of Psychology, 63*, 150-153. http://doi.org/10.1111/j.1742-9536.2011.00018.x

Hitchcock, D. (2018). The Stanford Encyclopedia of philosophy (Fall 2018 ed.). https://plato.stanford.edu/archives/fall2018/entries/critical-thinking/

Huber, C. R., & Kuncel N, R. (2016). Does college teach critical thinking? A metanalysis. *Review of Educational Research, 82*(2), 431-468. http://doi.org/10.3102/0034654315605917

Jones, J. H., & Morris, L. V. (2007). Evaluation of critical thinking skills in an associate degree nursing program. *Teaching and Learning in Nursing, 2,* 109-115. http://doi.org.10.1016/j.teln.2007.07.006

Kuhn, D., & Udell, W. (2003). The development of argument skills. *Child Development, 74*(5), 1245-1260. http://doi.10.1111/1467-8624.00605

Lawson, T. J. (1999). Assessing psychological critical thinking as a learning outcome of a Psychology majors. *Teaching of Psychology, 26*(3), 207-209. http:/doi.org/10.1207/S15328023TOP260311

Lawson, T. J., Jordan-Fleming, M. K., & Bodle, J. H. (2015). Measuring psychological critical thinking: An update. *Teaching of Psychology*, *42*(3), 248-253. http:/doi.org/10.1177/0098628315587624

Liu, O. L., Frankel, L., & Roohr, K. C. (2014). Assessing critical thinking in higher education: Current state and directions for next-generation assessment. *ETS Research Report Series 1*, 1-23.

Liu, O.L., Liu, H., Roohr, K. C., McCaffrey, D. F. (2016a). Investigating college learning gain: Exploring a propensity score weighting approach. *Journal of Educational Measurement*, *43*(3), 352-367). https://doi.org/10.1111/jedm.12112

Liu, O. L., Mao, L., Frankel, L., & Xu, J. (2016b). Assessing critical thinking in higher education: the HIEghtenTM approach and preliminary validity evidence. *Assessment & Evaluation in Higher Education*, 41(5), 677-694. http://doi.org/10.1080/02602938.2016.1168358

McLean, C. P., & Miller, N. A., (2010). Changes in critical thinking skills following a course on science and pseudoscience: A quasi-experimental study. *Teaching of Psychology, 37,* 85-90. https://doi.org/10.1080/00986281003626714

McPeck, J. E. (1981). *Critical thinking and education.* Robertson.

McPeck, J. E. (1990). Critical thinking and subject specificity: A reply to Ennis. *Educational Researcher, 19*(4), 10–12. http://doi.org/10.3102/0013189X019004010

Mines, R. A., King, P. M., Hood, A. B., & Wood, P. K. (1990). Stages of intellectual development and associated critical thinking skills in college students. *Journal of College Student Development, 31*(6), 538–547.

Musgrove, A. T., Powers, R. J., Rebar, L. C. & Musgrove, G. J. (2018) Real or fake? Resources for teaching college students how to identify fake news, *College & Undergraduate Libraries,* 25(3, 243-260. http://doi.org/10.1080/10691316.2018.1480444

O*Net Online (2019). *The occupational information network*. Retrieved September, 13, 2019, from https://www.onetonline.org/skills/

Ortiz, Ortiz, C. M. A. (2007). *Does philosophy improve critical thinking skills?* [Unpublished master's thesis]. The University of Melbourne, Victoria, Australia. https://www.reasoninglab.com/wp-content/uploads/2017/05/Alvarez-Final_Version.pdf

Pearson-LatentLens (2016). *Evaluating the admission process for a London based business school, April 2016. Pearson-TalentLens.*

Pascarella, E. T., & Terenzini, P. T. (2005). *How college affects students: A third decade of research*. Jossey-Bass.

Paul R. & Elder, L. (2019). *The thinker's guide to conscientious citizen on how to detect media bias and propaganda.* Rowman & Littlefield.

Renaud, R. D., & Murray, H. G., (2008). A comparison of subject-specific and a general measure of critical thinking. *Thinking Skills and Creativity, 3*(2), 85-93. https://doi.org/10.1016/j.tsc.2008.03.005

Pennycook, G., & Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition, 188*, 39-50. https://doi.org/10.1016/j.cognition.2018.06.011

Roohr, K. C., Liu, H., & Liu, O. L. (2017) Investigating student learning gains in college: a longitudinal study, *Studies in Higher Education*, *42*(12), 2284-2300. https://doi.org/10.1080/03075079.2016.1143925

Roohr, K., Olivera-Aguilar, M., Ling, G., & Rikoon, S. (2019). A multi-level modeling approach to investigating students' critical thinking at higher education institutions. *Assessment & Evaluation in Higher Education, 44*(2), 1-15. http://doi.org/10.1080/02602938.2018.1556776

Solon, T. (2007). Generic critical thinking infusion and course content learning in introduction to psychology. *Journal of Instructional Psychology, 34*(2), 95-109.

Stark, E. (2012). Enhancing and assessing critical thinking in a psychological research methods course. *Teaching of Psychology*, *39*(2), 107-112. https://doi.org/10.1177/0098628312437725

Steward, R., & Al-Abdulla, Y. (1989). An examination of the relationships between critical thinking and academic success on a university campus. *Research Report.* ERIC Report No. ED318936. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.843.7780&rep=rep1&type=pdf

The Bar Standards Board. (2015). *BCAT Impact and Performance Evaluation*. The Bar Standards Board.

The British Psychological Society (2019). Standards for the accreditation of undergraduate, conversion and integrated Masters programmes in psychology. Leicester. Retrieved September 12, 2019, from https://www.bps.org.uk/sites/bps.org.uk/files/Accreditation/Undergraduate%20Accreditation%20Handbook%202019.pdf

Wagner, T. A., & Harvey, R. J. (2006). Development of a new critical thinking test using item response theory. *Psychological Assessment, 18*(1), 100-5. http://doi.org/10.1037/1040-3590.18.1.100

Watson, G. B., & Glaser, E. M. (2002). *Watson-Glaser Critical Thinking Appraisal – UK Edition Manual*. The Psychological Corporation Limited.

Watson, G. B., & Glaser, E. M. (2018). *Watson-GlaserTM Critical Thinking Appraisal: User's guide and technical manual.* Pearson.

Wentworth, D. K., & Whitmarsh, L. (2017). Thinking like a psychologist introductory psychology writing assignments: Encouraging critical thinking and resisting plagiarism. *Teaching of Psychology, 44*(4) 335-341. http://doi.org/10.1177/0098628317727909

Williams, R., Oliver, R., & Stockdale, S. (2004). Psychological versus generic critical thinking as predictors and outcome measures in a large undergraduate human development course. *The Journal of General Education*, 53(1), 37-58. doi:10.1353/jge.2004.0022

Yanchar, S. C., Slife, B. D., & Warner, R. (2008). Critical thinking as disciplinary practice. *Review of General Psychology, 12*(3), 265-281. http://doi.org/10.1037/1089-2680.12.3.265

Appendix

Description and Examples of WGCTA Subtest

The examples provided here were retrieved from the following webpage

http://www.pearsonvue.com/phnro/wg_practice.pdf

Inference: An inference is a conclusion that a person can draw from certain observed or supposed facts.

Example

Statement: "Two hundred school students in their early teens voluntarily attended a recent weekend student conference in Leeds. At this conference, the topics of race relations and means of achieving lasting world peace were discussed, since these were problems that the students selected as being most vital in today's world."

Proposed Inference: As a group, the students who attended this conference showed a keener interest in broad social problems than do most other people in their early teens. (PT, because, as is common knowledge, most people in their early teens do not show so much serious concern with broad social problems. It cannot be considered definitely true from the facts given because these facts do not tell how much concern other young teenagers may have. It is also possible that some of the students volunteered to attend mainly because they wanted a weekend outing.) (Answer: TRUE; PROBABLY TRUE; INSUFFICIENT DATA; PROBABLY FALSE; FALSE).

Recognition of Assumptions: An assumption is something presupposed or taken for granted.

Example:

Statement: "We need to save time in getting there so we'd better go by plane."

Proposed assumption: Going by plane will take less time than going by some other means of transportation. (YES, it is assumed in the statement that the greater speed of a plane over the

speeds of other means of transportation will enable the group to reach its destination in less time.). (Answer: Yes or No)

Deduction: In this test, each exercise consists of several statements (premises) followed by several suggested conclusions.

Example:

Statement: "Some holidays are rainy. All rainy days are boring. Therefore**:**"

*Proposed conclusion***:** Some holidays are not boring. (NO, the conclusion does not follow, even though you may know that some holidays are very pleasant.) (Answer: Yes or No)

Interpretation: The task is to judge whether or not each of the proposed conclusions or generalizations are warranted beyond a reasonable doubt from the information in the statement.

Example

Statement: "A study of vocabulary growth in children from eight months to six years old shows that the size of spoken vocabulary increases from 0 words at age eight months to 2,562 words at age six years."

Proposed Conclusion: None of the children in this study had learned to talk by the age of six months. (YES, the conclusion follows beyond a reasonable doubt since, according to the statement, the size of the spoken vocabulary at eight months was 0 words.) (Answer: Yes or No).

Evaluation of Arguments: In making decisions about important questions, it is desirable to be able to distinguish between arguments that are strong and arguments that are weak, as far as the question at issue is concerned.

Example:

Statement: "Should all young people in the United Kingdom go on to higher education?"

Proposed arguments: Yes; college provides an opportunity for them to wear college scarves. (WEAK, this would be a silly reason for spending years in college.) (Answer: Strong Argument - Weak Argument).

For all the subtest, except for *Inference*, the candidate is required to make a binary choice for each item. Though these subtests differ in the way they measure a respondent's critical thinking they frequently require any subjective or personal-moral attitude towards the content of a particular item to be separated from its critical evaluation, in addition test takers need to be able to isolate the truth or falsity of a statement from whether it is logical or plausible.

Table 1

Pre-Qualifications of Participants, by Year Group.

Year of study/ Pre-qualifications	Year 1	Year 3
A-levels	75	82
HND/BTEC[a]	2	2
Access[b]	3	1
IB[c]	8	1
Combination	3	3
Other	3	1

[a]Higher National Diploma/Ordinary National Diploma, [b]Access to Higher Education Diploma, [c]International Baccalaureate.

Table 2

Means and Standard Deviations of Total WGCTA and Subtest for First and Third Year Students

Year of study/ (Sub)test score type[a]	**Year 1**		**Year 3**	
	M	*SD*	*M*	*SD*
Overall WGCTA[uk b]	122.64	51.67	150.36	55.19
Inference	5.30	3.19	6.67	3.02
Recognition of Assumptions	5.57	4.92	8.13	5.34
Deduction	4.72	4.41	5.75	3.52
Interpretation	7.89	3.85	8.80	3.95
Evaluation of Arguments	7.17	3.79	7.62	3.96

[a]Subtest scores adjusted by chance 0-16; [b]Scores adjusted by chance range 0-320.

Table 3

Results of the ANCOVA with Year 1 and Year 3 as the Fixed Factor and Age, A-Level Acores and DEPI Scores as Covariates

Scales	Main effect of year group F (p) η_p^2	Covariate: Age F (p) η_p^2	Covariate: A-Level scores F (p) η_p^2	Covariate: DEPI F (p) η_p^2
Inference	4.055 (.046)	1.301 (.256)	1.659 (.200)	0.189 (.665)
	0.026	0.009	0.011	0.001
Recognition of	4.467 (.036)	0.480 (.490)	4.226 (.042)	0.009 (.923)
Assumption	0.029	0.003	0.028	0.0001
Deduction	0.006 (.938)	1.252 (.265)	.168 (.683)	0.545 (.461)
	0.0001	0.001	0.001	0.004
Interpretation	0.439 (.509)	0.128 (.722)	3.803 (.053)	6.119 (.014)
	0.003	0.001	0.025	0.039
Evaluation of	0.559 (.456)	0 .374 (.542)	0.350 (.555)	0.076 (.783)
Arguments	0.004	0.003	0.002	0.001

Note. The degrees of freedom for each analysis was 1, 149.

Table 4

Means and Standard Deviations of Total WGCTA and Subtest for First and Third Year

Year of study/ Test score type[a]	**Year 1**		**Year 3**	
	M	*SD*	*M*	*SD*
Overall WGCTA[uk][b]	131.14	54.86	162.75	59.73
Inference	5.52	3.17	7.92	3.47
Recognition of Assumptions	6.19	5.06	9.24	4.68
Deduction	5.30	4.83	6.06	4.89
Interpretation	8.35	3.69	8.95	3.71
Evaluation of Arguments	7.43	3.58	8.60	3.96

[a]Subtest scores adjusted by change 0-16; [b]Scores adjusted by change range 0-320.

Table 5

Results of the Hierarchical Multiple Regression Analysis

Variable		***B***	***SE B***	***β***	**t**	**R^2**	**ΔR^2**
Step 1:	Constant	353.583	159.132		2.22*	.049	-.007
	Age	-11.843	7.758	-.215	-1.527		
	DEPI	-.781	6.286	-.017	-.124		
	A-Level	.616	.724	.120	.851		
Step 2							
	Deduction	6.934	.819	.546	8.462***	.685	.660
	Inference	8.799	1.173	.508	7.495***	.853	.838
	Recognition	4.407	.565	.358	7.797***	.935	.927
	Assumption						
	Evaluation	2.888	.505	.187	5.721***	.962	.956
	Arguments						
	Interpretation	3.644	.219	.233	16.618***	.995	.994

*$p < .05$. ***$p < .001$.