Journal of University Teaching & Learning Practice

Volume 18
Issue 6 *Standard issue 3*

Article 11

2021

Collaborative testing in Sport and Exercise Degrees: A comparison of first and third year students' perceptions

Jonathon Headrick
Griffith University, Australia, jonathon.headrick@griffith.edu.au

Brooke Harris-Reeves
Griffith University, Australia, b.harris-reeves@griffith.edu.au

Talei Daly-Olm
Griffith University, Australia, talei.daly-olm@alumni.griffithuni.edu.au

Recommended Citation

Headrick, J., Harris-Reeves, B., & Daly-Olm, T. (2021). Collaborative testing in Sport and Exercise Degrees: A comparison of first and third year students' perceptions. *Journal of University Teaching & Learning Practice, 18*(6), 155-173. https://doi.org/10.53761/1.18.6.11

Collaborative testing in Sport and Exercise Degrees: A comparison of first and third year students' perceptions

Abstract

Collaborative testing is recognised as an effective assessment approach linked to positive student outcomes including enhanced test performance and reduced assessment anxiety. While collaborative testing approaches appear beneficial to university students in general, it is unclear whether students from different year levels benefit to the same extent. Therefore, the overarching aim of this study was to compare the perceptions and performances of first and third- year undergraduate students taking part in collaborative testing on multiple occasions during a semester. It was predicted that first-year students would perceive the collaborative testing opportunities as more beneficial than third-years given their limited formative experiences with university assessment. Further, it was expected that students would generally perform at a higher level on collaborative versus individual tests in line with previous work. Student performance and perceptions of collaborative testing were collected on two occasions within a semester over a period of two years in both a first-year and third-year course. Quantitative and qualitative results revealed that first-year students were more receptive and perceived more benefits relating to collaborative testing than third-years despite the fact both cohorts generally performed at a higher standard on the collaborative versus individual components. These findings suggest that while collaborative testing is considered beneficial to most, if not all, students, the benefits appear to be greater for first-year student cohorts.

Practitioner Notes

1. Collaborative testing should be strategically implemented targeting students transitioning into university, or in their earlier years of study.
2. Incorporating collaborative testing approaches when designing first-year assessment may foster intrinsic motivation and enhance early university experiences.
3. Results suggest that collaborative testing can benefit academic performance in third-year students, but should be considered carefully as these students also reported higher pressure and tension.
4. Overall, collaborative testing approaches appear beneficial to all students, but are best suited to early year students to foster the development of university assessment skills.

Keywords

Active learning, collaborative testing, cooperative learning, intrinsic motivation, assessment

Introduction

The validity of assessment of learning is a long-debated issue, but the validity of assessment for learning is a newer area of dialogue. Assessment for learning is geared towards enhancing the learning processes experienced through assessment rather than only measuring learning outcomes from assessment performance (Taras, 2002; William, 2011). Here we report on the implementation of a collaborative testing approach designed to foster assessment for learning in the context of sport and exercise degrees at an Australian University. As detailed in the following sections, the specific aim was to determine whether perceptions of, and performance in collaborative testing varied between year level cohort groups. Based on these findings we propose key implications for the use of collaborative testing to inform how this innovative assessment approach might best be incorporated in higher education settings.

Literature Review

Collaborative Testing Background

Collaborative testing is an assessment approach where students are provided the opportunity to work in small groups during summative assessment items, such as examinations or quizzes (tests) (Helmericks, 1993; Meseke et al., 2010; Slusser & Erickson, 2006). Recent work has identified features of collaborative testing that are reflective of the assessment *for learning* approach including: increased understanding of content (Rao et al., 2002), developing the ability to work in teams (Chiocchio et al., 2016; McClelland et al., 2014), improved retention of content (Cortright et al., 2003; Kleinberg et al., 2018), along with embracing alternate perspectives and engendering the ability to analyse and interpret information (Falchikov & Thompson, 2008). Further work has also highlighted that group collaboration is demonstrably more valuable than the sum of individual efforts (Tracy, 2019). Group assignment or project work, particularly in higher education, focuses on this sum of individual efforts and contributions resulting in a combined end product to be assessed (Lejk & Wyvill, 2001). Collaborative testing differs from a group assignment or project work because the process and experience of individuals working collaboratively in a small group is central to the assessment, rather than focussing on the production of an end product (Vogler & Robinson, 2016). Given that individuals begin their life-long learning journey in school, refining the collaborative process during their tertiary education years through social interaction, communication and group processes (Markauskaite & Goodyear, 2017), it has become increasingly important for educators to investigate and understand the effect of collaboration within higher education settings (Tucker & Reynolds, 2006) and the impact this has on individual learners.

Ensuring that learners are afforded the opportunity to work collaboratively in a safe learning environment may better prepare these individuals before entering the workforce. Within higher education settings, collaboration is currently accomplished through a wide range of both formal and informal activities (Bloom, 2009; Helmericks, 1993). Collaborative testing is one activity built on the concept of cooperative learning and involves students working through assessments such as quizzes or examinations in dyads or small groups (Kapitanoff, 2009; Slusser & Erickson, 2006). Cooperative learning principles of positive interdependence and individual accountability are facilitated by incorporating collaborative approaches into assessment tasks (Kapitanoff, 2009). Collaborative assessment tasks provide an ideal platform for these principles, whereby students' collective efforts can produce individual achievement (positive interdependence) whilst their individual efforts are evaluated independently of each other (individual accountability) (Millis & Cottell Jr, 1997; Nam & Zellner, 2011; Slavin, 1996; Williams et al., 2005).

Considerations for Implementation

Collaborative testing as an assessment approach within higher education settings has yielded positive tangible outcomes, including enhanced student learning and academic performance (Bloom, 2009; Gilley & Clarkston, 2014; Mahoney & Harris-Reeves, 2017; Rao et al., 2002; Springer et al., 1999). Other evidence for the more subjective, but arguably equally important, outcomes of collaborative testing include reduced test anxiety (Eastridge & Benson, 2020), increased critical thinking and enhanced interpersonal skills (Kapitanoff, 2009; Lusk & Conklin, 2003; Wiggs, 2011). Collaborative testing has benefited students across a variety of disciplines ranging from the more creative-focused (e.g., Bloom, 2009; Tucker & Reynolds, 2006) through to the more applied (e.g., Mahoney & Harris-Reeves, 2017). These benefits exist in laboratory-based experiments, in real classrooms (LoGiudice et al., 2015) and across a range of questioning types (i.e., multiple-choice, short answer, true/false) (Kapitanoff, 2009; Rao et al., 2002).

While introducing collaborative approaches within assessment tasks has its advantages, there are some challenges that should also be noted. Both individual and group-level considerations exist when investigating students' experiences in collaborative approaches to learning and assessment. Differences in motivations, personalities and other group processes may influence students' experiences in collaborative testing and are difficult to control given the nature of individual differences. Individual level attributes include students' performance motivation (French & Kottke, 2013), personality traits, and socio-cultural background (Pineda et al., 2009; Tröster et al., 2014). Group level processes include social comparison, conformity, and conflict management (Ilarda & Findlay, 2006; Micari & Drane, 2011; Micari & Pazos, 2014; Mohammed & Angell, 2004).

Despite these limitations, qualitative accounts from students indicate that engaging in collaborative testing is largely an enjoyable experience (Mahoney & Harris-Reeves, 2019) with many students reporting an overall satisfaction with the level of discussion during the collaborative process (Cortright et al., 2003; Simkin & Kuechler, 2005; Zimbardo et al., 2003). These qualitative accounts provide complimentary and rich contextual information that helps to inform educators about students' perceptions of both the utility and experience of collaborative testing.

Who Benefits Most from Collaborative Testing?

As outlined above *the what* of collaborative testing has been well documented, including the benefits, limitations and rationale for implementing collaborative approaches to assessment in higher education settings. Whilst it is important to understand what collaborative testing is and its benefits, when the investigation shifts to focus on *who* benefits (most) from collaborative testing, the evidence is lacking. The contrast between high- and low-performing students is one example which examines who benefits from collaborative testing. Giuliodori and colleagues (2008) examined collaborative testing benefits among high- and low-performing students in an undergraduate veterinary physiology course and found collaborative testing was beneficial to performance for all students over and above individual testing, with the effects of collaborative testing of greatest benefit for the low-performing students. Mahoney and Harris-Reeves (2017) also found middle and lower performing students in an undergraduate sport and exercise psychology course increased their test results under collaborative test conditions compared to individual test conditions. Interestingly, upper performing students did not improve (nor worsen) in their overall individual and collaborative test results except when examining higher order thinking questions (HOTQ) specifically, in which all students regardless of academic ability (i.e., lower, middle and upper performers) benefitted equally from collaborative testing conditions compared to individual testing (Mahoney & Harris-Reeves, 2017). Although these previous studies have investigated the differential effects of collaborative versus individual testing across achievement level cohorts, the question remains whether other potential cohort differences exist. Taking this into consideration, the current study

aims to explore the utility of collaborative testing between year level cohorts. Specifically, the study set out to compare the perceptions and performances of first and third-year undergraduate students taking part in collaborative testing on multiple occasions throughout a semester.

We hypothesised that first-year students' would be more receptive to collaborative testing opportunities in terms of perceived value, competence, and interest metrics compared to their third-year counterparts. This is supported through the research of Taylor (2008), recognising carefully designed assessment can be a powerful tool that has the potential to assist first-year students transition into university. Peters, Jones, and Peters (2008) further support this premise through the conclusion that first and third-year exercise science students prefer quite distinct learning styles which has clear implications for how assessment (testing) might be designed for different year groups. We also predicted, that in line with previous collaborative testing investigations, students would generally perform at a higher level in the collaborative versus individual components of the assessment. Finally, it was expected that on subsequent collaborative testing opportunities (within the same semester), students would become more receptive and perform at a higher level as they develop collaborative strategies and recognise the potential academic benefits of this assessment style.

Method

Study Context

This study was embedded within two exercise and sport subjects consisting of equivalent contact hours per week in the form of a weekly 2-hour lecture and workshop over a 12-week teaching period (semester). The two selected in-class tests (per subject) were administered 6 weeks apart, were equally weighted, and assessed content from defined modules in the respective subjects. The assessment items in question involved an individual testing component where students responded to a set of multiple-choice questions (MCQs) and submitted their own answer sheet. The collaborative component provided the opportunity for students to discuss the same set of MCQs in small groups before each submitting a second answer sheet. Therefore, each student submitted two answer sheets per quiz contributing to their overall score for the subject.

Participants

Institutional ethical approval was obtained to collect survey responses from participants immediately after completing an individual and collaborative test (module-based quizzes) on two separate occasions during a 12-week semester. Participants were undergraduate Sport Development and Exercise Science students enrolled in either first-year or third-year subjects at an Australian University. Participants were recruited over two consecutive calendar years (2018-2019) to take part in a written survey following the two scheduled collaborative tests included in each subject. Of the prospective participant pool, 115 male (n = 69) and female (n = 46) students consented to complete the post-collaborative assessment survey on both occasions in their respective first-year (n = 73; Mean Age = 18.98 ± 2.14) or third-year (n = 42; Mean Age = 22.14 ± 3.45) subject. Participants identified as domestic (n = 104), international (n = 7), or indigenous students (n = 3) studying either full-time (n = 107) or part-time (n = 8). The majority of participants did not speak a language other than English (n = 99).

Students' Perception of Collaborative Testing

To ascertain perceptions of the collaborative testing method students were invited to respond to a twenty-three item version of the Intrinsic Motivation Inventory (IMI) and three open-ended questions (McAuley et al., 1989; Ryan, 1982; Ryan et al., 1990). The IMI is a validated (McAuley et al., 1989) and widely adopted instrument used to assess participant motivation relating to a

particular task with previous application in education contexts (Augustyniak et al., 2016; Cortright et al., 2013; Ryan et al., 1990). The version of the IMI (see Table 1) included five subscales, namely: Interest/Enjoyment (4 items); Perceived Competence (5 items); Effort/Importance (3 items); Pressure/Tension (5 items); and Value/Usefulness (6 items). Items relating to the Perceived Choice and Relatedness subscales of the full IMI were not included as they were deemed irrelevant to the task given that the students were required to complete the assessment as part of the course (i.e. no choice), and the collaborative testing experience itself was of interest, rather than the relationship with other students (relatedness). Participants were provided with verbal and written instructions detailing that the survey should be completed based on the "collaborative assessment you (i.e. the student) have just completed". The wording of the IMI items was contextualised to focus on the "assessment" wherever possible within the confines of the original item design (see Table 1). Items were presented in a randomised order and participants were asked to respond on a 7-point scale (1 = *Not at all True;* 7 = *Very True)*.

Three subsequent open-ended questions were included to provide participants with an opportunity to express more detailed perceptions or comments about the collaborative testing experience as follows:

1) Will your experience during this assessment piece change the way you approach similar assessment pieces in the future?
2) Do you believe that your capability to successfully complete similar assessment pieces has improved after completing this assessment?
3) Please provide any other comments or thoughts you have about this style of assessment.

Questions 1) and 2) were accompanied by check box options (Yes; No; Not Sure) and space for extra information to be provided regarding the reason(s) why/why not.

Test Performance

Performance in the respective assessment items was recorded for all participants based on the individual and collaborative component scores, along with relative changes. Percentage scores were calculated based on the correct responses selected out of the total available marks in both components to provide ease of comparison between quiz components, and cohorts (i.e. first vs third-year).

Procedure

Prospective participants were provided with the information regarding the purpose and nature of the study via class announcements and emails leading into each assessment occasion, along with written documentation provided with the survey. Students were informed that they were in no way obligated to take part in the study and their participation would not have any influence on their grades. Consent to participate was assumed by the return of a completed survey and participants were free to withdraw at any time.

Table 1.

Intrinsic Motivation Inventory (IMI) items and respective subscales included on the post assessment survey

Subscale	Items
Interest / Enjoyment	I would describe this assessment as very interesting.
	I thought this assessment was quite enjoyable.
	I thought this was a boring assessment. (R)
	This assessment did not hold my attention at all. (R)
Perceived Competence	After working at this assessment, I felt pretty competent.
	I think I did pretty well at this assessment, compared to other students.
	I am satisfied with my performance in this assessment.
	This was an assessment style that I couldn't do very well. (R)
	I think I am pretty good at this assessment style.
Effort / Importance	It was important to me to do well at this assessment.
	I tried very hard on this assessment.
	I didn't try very hard to do well at this assessment. (R)
Pressure / Tension	I was very relaxed in doing this assessment. (R)
	I did not feel nervous at all while doing this assessment. (R)
	I was anxious while working on this assessment.
	I felt very tense while doing this assessment.
	I felt pressured while doing is assessment.
Value / Usefulness	I believe this assessment could be of some value to me.
	I think doing this assessment could help me understand how to approach future assessment items.
	I think that doing this assessment is useful for developing confidence in myself.
	I believe doing this style of assessment is beneficial to me.
	I think this style of assessment is important to do because it can enhance my self-belief.
	I would be willing to do this style of assessment again because it has some value to me.

Note: (R) indicates items where indicated scores were reversed for analysis as per IMI instructions

Both the individual and collaborative tests components consisted of 30 minutes of working time, with the same 30 questions for each respective subject/module quiz, and required each student to submit their own responses on a separate response sheet for marking (Mahoney & Harris-Reeves, 2017). Following completion of the individual component all associated answer sheets were collected so that students could not alter their individual responses after collaboration. Students maintained possession of the question booklet and a second answer sheet to be used in the collaborative component. To allow collaboration students were instructed to form small working groups of their choice (maximum of 5 members). In these groups, students were offered the opportunity to discuss the same 30 MCQs and then submit a second set of individual responses. Following the return of all test materials students were invited to complete the paper-based survey reflecting on their experience with the collaborative testing before leaving the room. All copies (including blanks) of the survey were collected to avoid exposing students who had/had not completed the survey to the full cohort while also ensuring that any materials related to the assessment remained in the testing venue.

Data Analysis

Participant responses to the IMI and open-ended questions were collated alongside individual and collaborative component performance for each quiz occasion before being de-identified. Only responses from participants who completed the survey following both quizzes for each respective subject were included in the analysis. Subscale scores for the IMI were determined by producing averages across items corresponding to each subscale (Table 1). Quiz component performances were transformed to percentage scores for ease of comparison between individual – collaborative attempts, quiz occasions, and years.

To determine any high-level changes in overall quiz component (Individual vs Collaborative) performance (i.e., combined years, cohorts, and quiz numbers) one-way repeated measures ANOVAs were performed with quiz component as the within subject factor. To test for statistical differences in IMI subscale scores and (more in-depth) test performance a series of three-way (Year * Cohort * Quiz Number) mixed factorial ANOVAs were performed. The between-subject independent variables were 'Year' (2018 vs 2019) and 'Cohort' (first vs third-year), with 'Quiz Number' (Quiz 1 vs Quiz 2) as the within-subject independent variable given that all participants from respective years and cohorts completed both quizzes. All analyses were performed using IBM SPSS Statistics for Windows Version 26.0 (*IBM SPSS Statistics for Windows*, 2016) with alpha set at $p < .05$.

The open-ended questions were analysed utilizing a grounded theory methodology (Braun & Clarke, 2006). Braun and Clarke's (2006) six-phase process involve: familiarising with the data; generating initial codes; searching for themes; reviewing themes; defining and naming themes; and producing the paper. Responses to the qualitative questions were imported into MAXQDA (v. 18.2.0) Qualitative Software. MAXQDA was chosen as the qualitative software based on its multi-faceted functionality for thematic analysis. Following the familiarization with the texts, which included reading participant responses to the open questions, the transcripts were coded using a grounded theory approach to the analysis, allowing the data to "speak for themselves" without approaching the data with a pre-existing theoretical framework. The initial set of codes were developed manually, then a second set of codes were developed utilizing MAXQDA. Preliminary themes were identified and compared through a discussion of findings with the member of the research team. Codes were than grouped in themes, reviewed and defined. Themes were analysed in such a way to move from individual themes towards broad analytic themes. Additionally, Leximancer software was utilized to cross-check the manual and MAXQDA grounded theory output.

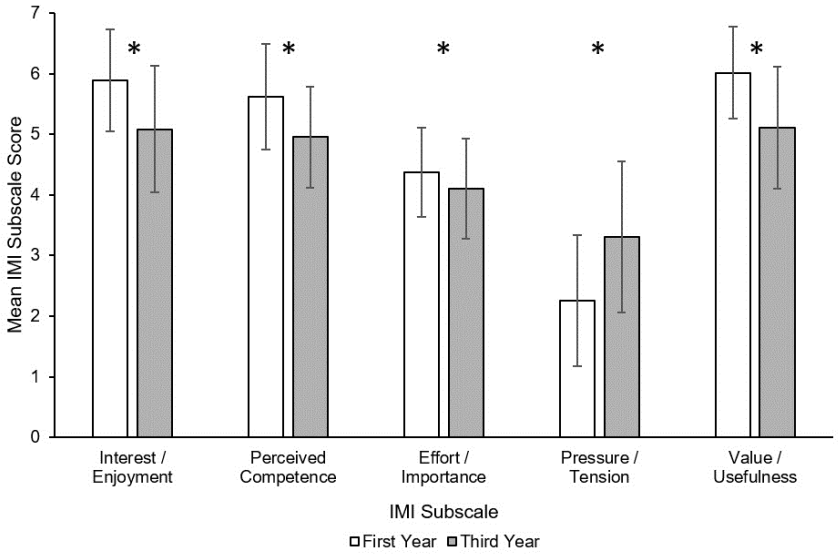
Results

IMI Subscales

As predicted, key differences were found in student perception of collaborative testing between first and third-year cohorts (between-subject) for all IMI subscales (Figure 1). Perceptions of Interest/Enjoyment ($F(1, 111) = 28.86$, $p < .001$), Competence ($F(1, 111) = 26.54$, $p < .001$), Effort/Importance ($F(1, 111) = 6.34$, $p < .05$), and Value/Usefulness ($F(1, 111) = 41.78$, $p < .001$) were all significantly higher for the first-year cohorts irrespective of quiz number or year (see Table 2 for mean and SDs). In contrast, the third-year cohorts reported higher perceptions of Pressure/Tension when compared with first-year cohorts ($F(1, 111) = 28.09$, $p < .001$). No statistical differences or interactions were observed comparing IMI subscales based on the factors of year (between-subject) or quiz number (within-subject) indicating that any differences in perception were cohort dependant.

Figure 1.

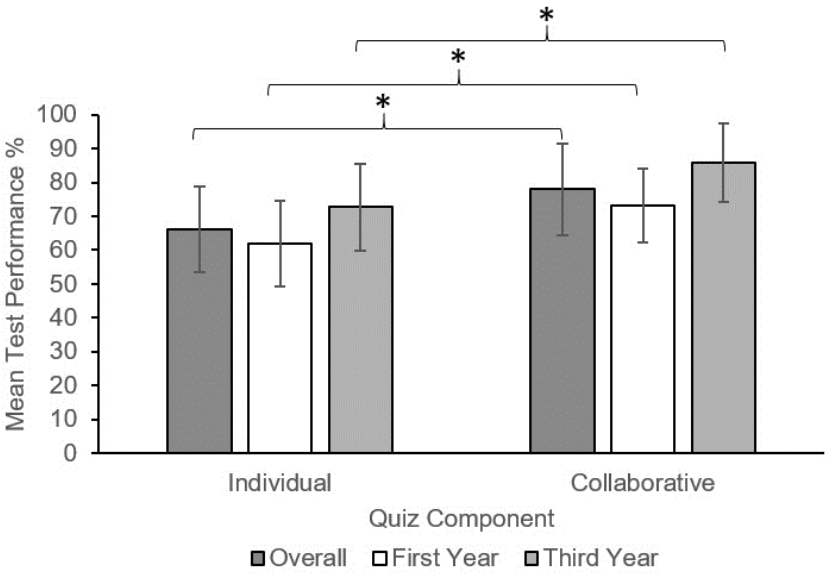
Mean IMI subscale scores for first and third-year participants irrespective of year or quiz number.



Note: Statistically significant differences between cohorts indicated by *****

Figure 2.

Mean individual and collaborative quiz performance scores (% out of 100) for first and third-year participants irrespective of year or quiz number



Note: Statistically significant differences between quiz components indicated by *

Test Performance

Overall comparisons of test performance regardless of year, cohort, or quiz number factors (Figure 2) indicated that student performances were significantly higher ($F(1, 223) = 311.40$, $p < .001$) in Collaborative ($M = 77.93$, $SD = 13.70$) versus Individual ($M = 66.05$, $SD = 12.72$) quiz components. Both first ($F(1, 139) = 174.07$, $p < .001$) and third-year ($F(1, 83) = 139.91$, $p < .001$) cohorts also revealed statistically significant test performance increases in the Collaborative (first-year $M = 73.19$, $SD = 10.94$; third-year $M = 85.83$, $SD = 11.53$) versus Individual (first-year $M = 62.01$, $SD = 12.6$; third-year $M = 72.78$, $SD = 12.86$) component.

More in-depth analyses of test performance revealed statistical differences and interactions based on year, cohort, and quiz number factors. Increases in Collaborative versus Individual component test performances were found to be statistically different between years ($F(1, 108) = 8.22$, $p < .01$) with scores from 2018 students increasing at a greater magnitude in the Collaborative component compared with 2019 students. A significant interaction was also observed for the quiz number * year relationship ($F(1, 108) = 5.52$, $p < .05$) indicating that Collaborative component scores improved in a different manner for respective quizzes across the two years. A significant main effect of cohort was found (between-subject; see Figure 2) for both the Individual ($F(1, 108) = 28.75$, $p < .001$) and Collaborative ($F(1, 108) = 54.52$, $p < .001$) components with the third-year cohorts tending to score higher in relation to their first-year counterparts on both quiz components.

Finally, a significant main effect of quiz number (within-subject) was observed for the Individual ($F(1, 108) = 16.31$, $p < .001$) and Collaborative ($F(1, 108) = 39.37$, $p < .001$) component scores indicating that participants generally (i.e., first and third-years combined) scored higher in quiz 1 compared with quiz 2. However, a significant interaction of quiz number * cohort was also found

for the Individual ($F(1, 108) = 33.67$, $p < .001$) and Collaborative ($F(1, 108) = 40.37$, $p < .001$) components indicating that the magnitude of change in performance from quiz 1 –to- quiz 2 was different between cohorts. Specifically, the performance of first-year students appeared to change (decrease) more substantially from quiz 1 – to – quiz 2 than that of third-years.

The qualitative data analysis yielded distinct themes from the open-questions around student perceptions of the collaborative testing. It is clear from the themes that first-year student experiences and perceptions of collaborative testing were different to those of the third-year students. The emerging themes from the first-year student data focused on 'understanding how to approach assessment', 'relaxed' and 'confidence'. While the themes that emerged from the third-year student data were 'collaborating with others' and 'approach'.

First-year themes

The first theme *understanding how to approach assessment* comprises of two categories - reading and answering exam questions and study techniques. The findings suggest that students found the process of talking through the answers and questions with fellow students to be beneficial. Specifically, it enabled them to understand how they were processing the questions and answers "it helped me to learn and identify my mistakes, while allowing me to fix them in part B" and "gives me a lot better understanding of why other points of view may be correct and gave me a better understanding of how I go about answering questions". In some cases students recognised that they were reading the questions incorrectly or second guessing their answers "it makes me realise not to second guess and change my answers", "it has taught me how to read questions differently" and "I will look at key words that I didn't pick up on in the first place when my friends did".

The second theme that emerged from the first-year data was *relaxed*. This theme refers to the reduced stress and anxiety experienced by students before, during and after the collaborative testing. The theme consisted of two categories – preparation and process. The students' experience of the collaborative testing resulted in an understanding of different strategies when preparing for the test "I am less stressed about assessment, I found I learnt more when talking to others about the content, so I plan to collaborate with peers when I study" and "I will feel more relaxed next time I prepare for the test (be)cause I now understand how to prepare". Students also expressed opinions about their feelings during the quiz, "I was far more relaxed and calm during this assessment because I was able to talk over the content with peers" and "the process of discussing the answers with others provided a great opportunity to critically evaluate each question and feel less stress".

The third theme, *confidence* highlighted how the student's self-belief was affected as a result of the collaborative approach. This was expressed in terms of similar assessment items "it gives you a chance to look back over your answers giving you a feeling of self-belief" and "I believe that my confidence will increase, and I won't second guess my final answer in the next test". Students also reported an increase in their ability to successfully prepare for and complete future assessment items "feeling confident goes a long way to positive results, I felt extremely confident. I will now have more confidence in myself while completing assessments in the future" and "this collaborative assessment has given me confidence that I am doing MCQ correctly and it means I will have confidence in other subjects".

Table 2.

Mean (SD) Intrinsic Motivation Inventory (IMI) subscale results for quiz number, cohort, and year.

	Quiz 1				Quiz 2			
IMI Subscale	Cohort	Mean (SD)	Year	Mean (SD)	Cohort	Mean (SD)	Year	Mean (SD)
Interest / Enjoyment	1st Year	5.88 (0.88)	2018	5.98 (0.91)	1st Year	5.90 (0.80)	2018	6.00 (0.76)
			2019	5.81 (0.86)			2019	5.83 (0.83)
	3rd Year	5.15 (1.01)	2018	5.19 (0.86)	3rd Year	5.02 (1.09)	2018	5.04 (1.13)
			2019	5.11 (1.16)			2019	5.00 (1.07)
Perceived Competence	1st Year	5.65 (0.79)	2018	5.67 (0.65)	1st Year	5.60 (0.95)	2018	5.72 (0.88)
			2019	5.64 (0.89)			2019	5.50 (0.99)
	3rd Year	4.88 (0.77)	2018	4.71 (0.77)	3rd Year	5.03 (0.88)	2018	4.85 (1.04)
			2019	5.05 (0.75)			2019	5.21 (0.67)
Effort / Importance	1st Year	4.44 (0.70)	2018	4.60 (0.65)	1st Year	4.32 (0.77)	2018	4.52 (0.78)
			2019	4.32 (0.72)			2019	4.17 (0.74)
	3rd Year	4.08 (0.72)	2018	4.00 (0.67)	3rd Year	4.13 (0.92)	2018	4.11 (1.14)
			2019	4.15 (0.78)			2019	4.15 (0.66)
Pressure / Tension	1st Year	2.27 (1.06)	2018	2.17 (1.04)	1st Year	2.25 (1.10)	2018	2.26 (0.91)
			2019	2.33 (1.07)			2019	2.23 (1.24)
	3rd Year	3.33 (1.15)	2018	3.18 (1.14)	3rd Year	3.29 (1.35)	2018	3.39 (1.32)
			2019	3.48 (1.17)			2019	3.18 (1.40)
Value / Usefulness	1st Year	6.11 (0.74)	2018	6.10 (0.71)	1st Year	5.93 (0.77)	2018	5.99 (0.80)
			2019	6.12 (0.76)			2019	5.88 (0.75)
	3rd Year	5.12 (0.90)	2018	5.15 (0.83)	3rd Year	5.11 (1.12)	2018	5.10 (1.03)
			2019	5.08 (0.98)			2019	5.12 (1.22)

Third-year themes

The central theme in the third-year student data was *collaborating with others*. The data suggests that students fell into two distinct groups: collaborative testing as a positive experience and collaborative testing as a stress inducing, negative experience. When reporting on their experiences, many students explained they benefited from the interactions "I really enjoy collaborative exams because I believe you learn more hearing what others have to say" and "it was a great experience to work with others and develop different skills". On the contrary, some did not find it to be beneficial stating "seeing how others work through things put doubt in my mind", "I like my own way of doing things" and "listening to others makes me second guess myself, I'd rather do it alone".

Discussion

The intention of this study was to explore student perceptions of collaborative testing, with a focus on the perceptions of first-year versus third-year cohorts over two successive years. We hypothesised that students from first-year cohorts would be more supportive and perceive greater benefit from collaborative testing opportunities given their limited exposure to university level education and assessment. As detailed in the previous section, this hypothesis was supported by the data with first-year student responses indicating higher perceptions of value/usefulness, interest/enjoyment, perceived competence, and effort/importance in the collaborative testing process. In contrast, third-year students reported higher levels of pressure/tension arising from the collaborative testing experience. These trends were evidenced by responses to the IMI and supported by additional open-ended responses offering insights into the student experience.

In relation to our second hypothesis, results indicated that, on average, all students performed at a higher level in the collaborative quiz components in line with previous work. Interestingly, and contrary to our third hypothesis, students did not appear to be more receptive, nor perform at a higher level during the second collaborative testing occasion. In fact, overall test performances were found to decrease from quiz 1 – to quiz 2 with differing trends evident based on cohort and year. In summary, all students appeared to benefit in terms of test performance following collaborative quiz components. However, there is a case to be made that benefits for first-year students are more pronounced considering the differences in participant perceptions between cohorts.

Student perceptions of collaborative testing

Student perception based on IMI subscale ratings for the two cohort groups reflected similar overall trends (e.g., value subscale rating highest and pressure subscale lowest). These findings echo previous work concluding that collaborative testing is generally beneficial in terms of fostering student engagement and developing assessment skills (Kapitanoff, 2009). Critically, our findings revealed marked differences between the first and third-year groups indicating that, while beneficial to all, collaborative testing activities appear to have the greatest impact for students in the early stages of their university study. This finding was consistent, irrespective of the cohort that was sampled and the separate quiz occasions within the two sample cohorts. Results from the subscales of value/usefulness and pressure/tension were perhaps most noteworthy in highlighting the differences in perceptions between cohorts. Consistent with previous research findings (Kapitanoff, 2009; Lusk & Conklin, 2003; Wiggs, 2011) these results were supported by the qualitative data where first-year students reported reduced stress/anxiety, and consequently tended to feel more relaxed during the collaborative testing. These findings align with previous work recognising that assessment items should be strategically tailored to meet the needs of first-year students to foster their transition into university and the development of university level assessment skills (Taylor, 2008). Year group differences in preferred learning styles have also been identified between first and third-year students in sport-related degrees (e.g., exercise science, sport studies, coaching, and

leisure studies) that are congruent with the students in the current study (Peters et al., 2008). Furthermore, as indicated by Wagner et al., (2014) students in exercise science degrees have a strong tendency to develop multiple learning styles throughout their studies which further supports the implementation of collaborative testing in early years of university to afford exploration of learning and test completion strategies in group settings. Therefore, we suggest that a collaborative testing approach is most effective for students transitioning into university studies and provides an opportunity to develop academic skills as independent learners (i.e., assessment for learning).

Student test performance

In line with previous collaborative testing research from wider higher education settings (Bloom, 2009; Lusk & Conklin, 2003), the broader sciences (Gilley & Clarkston, 2014; Rao et al., 2002; Springer et al., 1999), and more specifically exercise and sport degrees (Cortright et al., 2003; Mahoney & Harris-Reeves, 2017) improvements in performance were observed between the individual and collaborative components for all quizzes. One objective of this study was to identify the difference, if any, in test performances for first-year and third-year cohorts. Results revealed that third-years tended to achieve higher scores (~10%) than first-years for both individual and collaborative components. Interestingly, on average both cohorts improved their respective scores in the collaborative components at similar rates, with statistical differences in change related to quiz number and year only. Therefore, it can be concluded that the opportunity to be involved in collaborative testing benefited student performance (overall) at similar magnitudes for different cohort groups, but somewhat fluctuated based on quiz number and calendar year. Such fluctuations are expected given the different content assessed on each quiz and modifications to quiz questions made year-to-year.

Practical Implications

Drawing on the key findings detailed above, this study provides a novel contribution to existing work involving collaborative testing. While the current paper focusses on students studying undergraduate Exercise Science and Sport Development degrees, similar findings from other disciplines and study areas suggest that the approach presented is a viable option for developing skills for life-long learning. Specifically, we propose that while collaborative testing is beneficial (perceptions and performance) to all students it is particularly worthwhile for implementation in cohorts of students who are transitioning into university. Critically, the collaborative testing experience was held in higher value by first-year students in terms of opportunities to develop successful assessment strategies in a group setting early in their degree. Alternatively, third-year students perceived less value and also reported higher perceptions of pressure / tension during the collaborative process compared with first-year students. Therefore, we suggest that collaborative testing should be implemented with caution in later year university cohorts, particularly those who have not experienced collaborative testing in earlier years of study. However, for first-year students, we advocate for collaborative testing to be strategically integrated into assessment plans to capitalise early on the raft of benefits to student capability, satisfaction, and performance in university learning and assessment.

Considering the current and evolving constraints on university learning and teaching practice imposed by the novel coronavirus pandemic (COVID-19) the above implications must also be placed in context (Green et al., 2020). This study takes the form of a "study before" the outbreak of COVID-19 describing methods and reporting on data with no direct relevance to addressing issues presented by the current pandemic (Crawford, 2021). Despite major changes to learning and teaching practices in response to the pandemic around the globe (see Crawford et al., 2020) the key findings and conclusions presented in this study remain relevant and have the potential to impact the university experience of students in the future. Collaborative testing remains a viable assessment

(and learning) strategy either during on campus classes (adhering to local social distancing restrictions) or through the use of virtual platforms offering synchronous "break out" groups or rooms. Indeed, collaborative testing may offer students the opportunity to interact and communicate with their peers during the (often forced) transition away from face-to-face and/or synchronous learning experiences (Crawford et al., 2020; Green et al., 2020). It is worth further noting that COVID-19 has had an impact on the collective well-being of many and higher education is not exempt from the pressures that COVID-19 can add to the learner experience, particularly for first-time university students (see Burns et al., 2020). Therefore, the current study's findings further illustrate that collaborative testing as an assessment approach has the potential to enhance the learner experience (particularly for first-year students) and aide the transition to university which may look and feel unconventional in the post COVID-19 environment.

Limitations and Future Directions

While our findings support the value of collaborative testing in the first-year of university, it is also important to highlight the potential limitations of this research while raising important questions for future work. First, the exam structure of this collaborative testing was only multiple-choice items. This raises the question, what effect does collaborative testing have on first-year student perceptions in different types of examination structures (e.g. short answer and essay style responses) that promote deeper student learning? This would build on previous research (Mahoney and Harris-Reeves, 2019) that explored the effects of collaborative testing on higher order thinking in third-year students. Second, students were only exposed to collaborative testing on two occasions. Therefore, future research could explore the effect of repeated exposure, and if there is a shift in student perceptions when exposed to repeated collaborative testing.

Conclusions

Higher education institutions are responsible for providing curriculum and assessment that enhances student learning, which adequately prepares students to enter the workplace as capable professionals. Our findings highlight the importance of introducing collaborative testing specifically to first-year university students. Whilst test performance improved for both first and third-year students during collaborative testing, first-years reported enhanced perceived value, interest, and competence in collaborative testing opportunities compared with their third-year counterparts. Open-ended questioning also provided rich contextual information that indicated first-year students viewed collaborative testing as an opportunity to build academic capability as they learn how to approach assessment in the formative years of university education. Despite benefitting academically from collaborative testing, third-year students appear to be less receptive and reported higher scores on the pressure/tension subscale of the IMI. In summary, our findings provide further support that collaborative testing is generally beneficial to overall student performance and satisfaction. However, it appears particularly beneficial for first-year students as an assessment approach to facilitate their initial development in becoming life-long learners.

References

Augustyniak, R. A., Ables, A. Z., Guilford, P., Lujan, H. L., Cortright, R. N., & DiCarlo, S. E. (2016). Intrinsic motivation: an overlooked component for student success. *Advances in Physiology Education, 40*(1), 465-466. https://doi.org/10.1152/advan.00072.2016.

Bloom, D. (2009). Collaborative Test Taking: Benefits for Learning and Retention. *College Teaching, 57*(1), 216-220. https://doi.org/10.1080/87567550903218646

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Burns, D., Dagnall, N., & Holt, M. (2020). Assessing the impact of the COVID-19 pandemic on student wellbeing at Universities in the United Kingdom: A conceptual analysis. *Frontiers in Education, 5*(582882). https://doi.org/10.3389/feduc.2020.582882

Chiocchio, F., Lebel, P., & Dubé, J. (2016). Informational role self-efficacy: A validation in interprofessional collaboration contexts involving healthcare service and project teams. *BMC Health Services Research, 16*(1), 153. https://doi.org/10.1186/s12913-016-1382-x

Cortright, R. N., Collins, H. L., Rodenbaugh, D. W., & DiCarlo, S. E. (2003). Student retention of course content is improved by collaborative-group testing. *Advances in Physiology Education, 27*(3), 102-108. https://doi.org/10.1152/advan.00041.2002

Cortright, R. N., Lujan, H. L., Blumberg, A. J., Cox, J. H., & DiCarlo, S. E. (2013). Higher levels of intrinsic motivation are related to higher levels of class performance for male but not female students. *Advances in Physiology Education, 37*(1), 227-232. https://doi.org/10.1152/advan.00018.2013.

Crawford, J. (2021). During and beyond a pandemic: Publishing learning and teaching research through COVID-19. *Journal of University Teaching & Learning Practice, 18*(3), Art. 02. https://ro.uow.edu.au/jutlp/vol18/iss3/02

Crawford, J., Butler-Henderson, K., Rudolph, J., Malkawi, B., Glowatz, M., Burton, R., Magni, P. A., & Lam, S. (2020). COVID-19: 20 countries' higher education intra-period digital pedagogy responses. *Journal of Applied Learning & Teaching, 3*(1), 9-28. https://doi.org/10.37074/jalt.2020.3.1.7

Eastridge, J. A., & Benson, W. L. (2020). Comparing two models of collaborative testing for teaching statistics. *Teaching of Psychology, 41*(1), 68-73. https://doi.org/10.1177/0098628319888113

Falchikov, N., & Thompson, K. (2008). Assessment: What drives innovation? *Journal of University Teaching & Learning Practice, 5*(1), 49-60. https://doi.org/https://ro.uow.edu.au/jutlp/vol5/iss1/5

French, K. A., & Kottke, J. L. (2013). Teamwork satisfaction: Exploring the multilevel interaction of teamwork interest and group extraversion. *Active Learning in Higher Education, 14*(3), 189-200. https://doi.org/10.1177/1469787413498034

Gilley, B. H., & Clarkston, B. (2014). Collaborative testing: Evidence of learning in a controlled in-class study of undergraduate students. *Journal of College Science Teaching, 43*(3), 83-91. https://doi.org/jstor.org/stable/43632038

Giuliodori, M. J., Lujan, H. L., & DiCarlo, S. E. (2008). Collaborative group testing benefits high- and low-performing students. *Advances in Physiology Education, 32*(4), 274-278. https://doi.org/10.1152/advan.00101.2007.

Green, W., Anderson, V., Tait, K., & Tran, L. T. (2020). Precarity, fear and hope: reflecting and imagining in higher education during a global pandemic. *Higher Education Research & Development, 39*(7), 1309-1312. https://doi.org/10.1080/07294360.2020.1826029

Helmericks, S. G. (1993). Collaborative testing in social statistics: Toward Gemeinstat. *Teaching Sociology, 21*(3), 287-297. https://doi.org/10.2307/1319027

IBM SPSS Statistics for Windows. In. (2016). (Version 24.0) IBM Corp.

Ilarda, E., & Findlay, B. M. (2006). Emotional intelligence and propensity to be a teamplayer *E-Journal of Applied Psychology: Emotional Intelligence, 2*(2), 19-29.

Kapitanoff, S. H. (2009). Collaborative testing: Cognitive and interpersonal processes related to enhanced test performance. *Active Learning in Higher Education, 10*(1), 56-70. https://doi.org/10.1177/1469787408100195

Kleinberg, K., Eastwood, J. L., & Rodenbaugh, D. W. (2018). Efficacy of collaborative testing for long-term retention of medical knowledge. *The Journal of the Federation of American Societies for Experimental Biology, 32*, 21.26. https://doi.org/10.1096/fasebj.2018.32.1_supplement.21.6

Lejk, M., & Wyvill, M. (2001). The effect of the inclusion of self-assessment with peer assessment of contributions to a group project: a quantitative study of secret and agreed assessments. *Assessment & Evaluation in Higher Education, 26*(6), 551-561. https://doi.org/10.1080/02602930120093887

LoGiudice, A. B., Pachai, A. A., & Kim, J. A. (2015). Testing together: When do students learn more through collaborative tests? *Scholarship of Teaching and Learning in Psychology, 1*(4), 377-389. https://doi.org/10.1037/stl0000041

Lusk, M., & Conklin, L. (2003). Collaborative testing to promote learning. *Journal of Nursing Education, 42*(3), 121-124. https://doi.org/10.3928/0148-4834-20030301-07

Mahoney, J., & Harris-Reeves, B. (2017). The effects of collaborative testing on higher order thinking: Do the bright get brighter? *Active Learning in Higher Education*, 1-13. https://doi.org/10.1177/1469787417723243

Markauskaite, L., & Goodyear, P. (2017). Epistemic fluency and professional education: Innovation, knowledgeable action and actionable knowledge. Springer. https://doi.org/10.1007/978-94-007-4369-4

McAuley, E., Duncan, T., & Tammen, V. V. (1989). Psychometric properties of the intrinsic motivation inventory in a competitive sport setting: A confirmatory factor analysis. *Research Quarterly for Exercise and Sport, 60*, 48-58.

McClelland, G. P., Leach, D. J., Clegg, C. W., & McGowan, I. (2014). Collaborative crafting in call centre teams. *Journal of Occupational and Organizational Psychology, 87*(3), 464-486. https://doi.org/10.1111/joop.12058

Meseke, C. A., Nafziger, R., & Meseke, J. K. (2010). Student attitudes, satisfaction, and learning in a collaborative testing environment. *The Journal of Chiropractic Education, 24*(1), 19-29. https://doi.org/10.7899/1042-5055-24.1.19

Micari, M., & Drane, D. (2011). Intimidation in small learning groups: The roles of social-comparison concern, comfort, and individual characteristics in student academic outcomes. *Active Learning in Higher Education, 12*(3), 175-187. https://doi.org/10.1177/1469787411415078

Micari, M., & Pazos, P. (2014). Worrying about what others think: A social-comparison concern intervention in small learning groups. *Active Learning in Higher Education, 15*(3), 249-262. https://doi.org/10.1177/1469787414544874

Millis, B. J., & Cottell Jr, P. G. (1997). *Cooperative Learning for Higher Education Faculty*. ABC-CLIO.

Mohammed, S., & Angell, L. C. (2004). Surface- and deep-level diversity in workgroups: Examining the moderating effects of team orientation and team process on relationship conflict. *Journal of Organizational Behavior, 25*(8), 1015-1039. https://doi.org/10.1002/job.293

Nam, C. W., & Zellner, R. D. (2011). The relative effects of positive interdependence and group processing on student achievement and attitude in online cooperative learning. *Computers & Education, 56*(3), 680-688. https://doi.org/10.1016/j.compedu.2010.10.010

Peters, D., Jones, G., & Peters, J. (2008). Preferred 'learning styles' in students studying sports-related programmes in higher education in the United Kingdom. *Studies in Higher Education, 33*(2), 155-166. https://doi.org/10.1080/03075070801916005

Pineda, R. C., Barger, B., & Lerner, L. D. (2009). Exploring differences in student perceptions of teamwork: The case of U.S. and Lithuanian students. *Journal of International Business and Cultural Studies, 1*, 50-58.

Rao, S. P., Collins, H. L., & DiCarlo, S. E. (2002). Collaborative testing enhances student learning. *Advances in Physiology Education, 26*(1), 37-41. https://doi.org/10.1152/advan.00032.2001

Ryan, R. M. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of Personality and Social Psychology, 43*(3), 450-461. https://doi.org/10.1037/0022-3514.43.3.450

Ryan, R. M., Connell, J. P., & Plant, R. W. (1990). Emotions in nondirected text learning. *Learning and Individual Differences, 2*(1), 1-17. https://doi.org/10.1016/1041-6080(90)90014-8

Simkin, M. G., & Kuechler, W. L. (2005). Multiple-Choice Tests and Student Understanding: What Is the Connection? *Decision Sciences, 3*(1), 73-98. https://doi.org/10.1111/j.1540-4609.2005.00053.x

Slavin, R. E. (1996). Research on cooperative learning and achievement: What we know, what we need to know. *Contemporary Educational Psychology, 21*(1), 43-69. https://doi.org/10.1006/ceps.1996.0004

Slusser, S. R., & Erickson, R. J. (2006). Group quizzes: An extension of the collaborative learning process. *Teaching Sociology, 34*(1), 249-262. https://doi.org/10.1177/0092055X0603400304

Springer, L., Stanne, M. E., & Donovan, S. S. (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis. *Review of Educational Research, 69*(1), 21-51. https://doi.org/10.3102/00346543069001021

Taras, M. (2002). Using assessment for learning and learning from assessment. *Assessment & Evaluation in Higher Education, 27*(6), 501-510. https://doi.org/10.1080/0260293022000020273

Taylor, J. A. (2008). Assessment in first year University: A model to manage transition. *Journal of University Teaching & Learning Practice, 5*(1), 19-33. https://ro.uow.edu.au/jutlp/vol5/iss1/3

Tracy, M. (2019). Lessons from MARS: an interview with Carlos Valdes Dapena. *Development and Learning in Organizations, 33*(1), 37-38. https://doi.org/10.1108/DLO-01-2019-169

Tröster, C., Mehra, A., & van Knippenberg, D. (2014). Structuring for team success: The interactive effects of network structure and cultural diversity on team potency and performance. *Organizational Behavior and Human Decision Processes, 124*(2), 245-255. https://doi.org/10.1016/j.obhdp.2014.04.003

Tucker, R., & Reynolds, C. (2006). The Impact of Teaching Models, Group Structures and Assessment Modes on Cooperative Learning in the Student Design Studio. *Journal for Education and the Built Environment, 1*(2), 39-56. https://doi.org/10.11120/jebe.2006.01020039

Vogler, J. S., & Robinson, D. H. (2016). Team-based testing improves individual learning. *The Journal of Experimental Education, 84*(4), 787-803. https://doi.org/10.1080/00220973.2015.1134420

Wagner, M. G., Hansen, P., Rhee, Y., Brunt, A., Terbizan, D., & Christensen, B. (2014). Learning style preferences of undergraduate Dietetics, Athletic Training, and Exercise Science students. *Journal of Education and Training Studies, 2*(3), 198-205. https://doi.org/10.11114/jets.v2i2.331

Wiggs, C. M. (2011). Collaborative testing: Assessing teamwork and critical thinking behaviors in baccalaureate nursing students. *Nurse Education Today, 31*(3), 279-282. https://doi.org/10.1016/j.nedt.2010.10.027

William, D. (2011). What is assessment for learning. *Studies in Educational Evaluation, 37*, 3-14. https://doi.org/10.1016/j.stueduc.2011.03.001

Williams, R. L., Carroll, E., & Hautau, B. (2005). Individual Accountability in Cooperative Learning Groups at the College Level: Differential Effects on High, Average, and Low Exam Performers. *Journal of Behavioral Education, 14*(3), 167-188. https://doi.org/10.1007/s10864-005-6296-3

Zimbardo, P. G., Butler, L. D., & Wolfe, V. A. (2003). Cooperative college examinations: More gain, less pain when students share information and grades. *The Journal of Experimental Education, 71*(2), 101-125. https://doi.org/doi.org/10.1080/00220970309602059