How well do students know their writing? A pilot study examining authorship quizzes to promote learning and address contract cheating

Matthew Quesnel, Brenda M. Stoesz
The Centre for The Advancement of Teaching and Learning, University of Manitoba

Contract cheating is a significant concern in higher education, requiring a multifaceted teaching and learning approach to address it. Quizzing students about their writing to promote engagement, confirm authorship and detect cheating has not yet been investigated systematically. Therefore, in this study, our objective was to explore the validity and usability of the quizzing software Auth+ to verify assessment authorship and detect contract cheating. A total of 24 third-year computer science students submitted their assignments to Auth+ and were quizzed, then completed a working memory assessment and a user experience survey. Students reported that Auth+ was easy to use and would discourage cheating but increased anxiety and contributed little to their learning. Quiz scores were correlated with individual differences in working memory capacity but not with other contract cheating indicators. These findings suggest that the quizzing method may have value, but anxiety and memory issues could be barriers to the usefulness of the software. In its current state, the effectiveness of Auth+ in verifying authorship and detecting contract cheating is limited. Further software development is required to increase quiz score validity as an indicator of outsourced academic work before higher education instructors or institutions adopt it as part of an academic integrity framework.

Implications for practice or policy:
- Authorship verification software that quizzes students on the content and style of their assignments show promise for deterring contract cheating and upholding academic integrity.
- Educators and institutions should not, however, rely on a single source of evidence to detect academic misconduct.
- Testing students on their writing using technology is similar to viva voce but it can scale up this learning and assessment process for large class sizes.

Keywords: academic integrity, authorship verification, detecting contract cheating, educational technology, working memory

Introduction

Contract cheating, or the act of engaging a third-party to complete academic work on one's behalf (Clarke & Lancaster, 2006), is a significant problem for students, instructors and educational institutions. Outsourcing academic work limits student engagement in course content and assessments, can lead to failure, suspension, or expulsion and threatens the validity of assessments and the credentials awarded. Self-reported data collected to estimate the prevalence of contract cheating suggests that a meaningful proportion (~ 2%–4%) of postsecondary students engage in this behaviour (Bretag, Harper, Burton, Ellis, Newton, Rozenberg et al., 2019). Newer research methods that incentivise honesty and correct for underreporting point to a higher prevalence (~ 8%–11.5%; Curtis et al., 2022). Contract cheating is difficult to detect because the work is typically bespoke. This sets it apart from plagiarism, as there is no original source to compare the work against. As such, a multifaceted approach that includes a focus on student learning as well as deterrence and detection is necessary to address contract cheating in higher education (Bertram Gallant, 2008, 2017; Bertram Gallant & Drinan, 2008; Bretag, Harper, Burton, Ellis, Newton, Rozenberg et al., 2019).

Technology that encourages students to engage in learning by answering quiz questions about their submitted text-based assignments could be an important component of a multifaceted approach to addressing contract cheating. To our knowledge, this approach has yet to be investigated systematically and is what we explored in this study. Our goal was to gain a preliminary understanding of students' experiences with an authorship verification software and obtain their perceptions of its value for learning and as an academic misconduct deterrent. We examined whether the quiz scores generated by the software were associated with individual differences in working memory or metacognitive self-regulation strategies (Kosnin, 2007; Pintrich et al., 1993). We also explored the relationship of authorship quiz scores with assignment grades and other indicators of outsourced academic work (Rogerson, 2017) to examine the validity of the technology for contract cheating detection. Software that purports to detect contract cheating is likely to be of great interest to educational institutions and educators looking for ways to ensure students' work is completed with integrity; thus, it is important to establish the validity of such software.

Contract cheating

Contract cheating can be categorised as non-commercial (i.e., outsourcing academic work to friends, family, peers, or tutors) or commercial (i.e., hiring essay writing and copyediting services to complete assignments; Lancaster & Clarke, 2016). With recent advances in generative artificial intelligence (GenAI), students can also use large language models to generate and complete a variety of assessment types (Cotton et al., 2023). When the offloading of academic work using GenAI is unauthorised and undeclared, the act may be classified as contract cheating, specifically, or academic misconduct more broadly (Foltynek et al., 2023). Academic integrity experts recommend implementing good teaching practices (e.g., constructive alignment, clear communication about course and program expectations) to encourage students to participate fully in the learning process and refrain from engaging in academic misconduct (Bretag, Harper, Burton, Ellis, Newton, Rozenberg et al., 2019; Bretag, Harper, Burton, Ellis, Newton, van Haeringen et al., 2019; Ellis et al., 2020; Harper et al., 2021). Good teaching practices, however, are unlikely to eliminate cheating entirely. Even assessments based on recommended strategies for deterring contract cheating (e.g., authentic assessment) can be outsourced (Bretag, Harper, Burton, Ellis, Newton, van Haeringen et al., 2019; Ellis et al., 2020). This highlights the need for a multifaceted approach to promote academic integrity and prevent academic misconduct, of which software that queries students about their writing may be one aspect.

Guidance for detection is available, but these practices can be labour and time intensive as they often involve manual inspection of students' academic work (e.g., Alin, 2020; Dawson & Sutherland-Smith, 2019; Rogerson, 2017; Tertiary Education Quality and Standards Agency [TEQSA], n.d.). Identifying academic misconduct through time-consuming manual inspection may reduce the motivation and willingness of instructors to search for evidence and report cheating (Eaton et al., 2020; Hamilton & Wolsky, 2022). Software that automates comparisons of works based on stylometry (often used in non-academic settings) notes similarities and differences in linguistic elements (e.g., word choice, sentence length, complexity) to determine authorship (Dawson et al., 2020; Ison, 2020; Juola, 2017). Both methods of examining student work for evidence of outsourcing require the availability of exemplars of students' prior writing, and those of other students for comparison, which may not be feasible. Moreover, if some or all of a student's submitted academic work was completed by the same third party or produced by GenAI, academic misconduct may not be suspected (Lines, 2016).

Authorship verification

In higher education, technology is often sought after to engage learners, deter inappropriate choices and streamline tasks associated with the practice of teaching. Auth+, an authorship verification software, claims to meet these educational goals by quizzing students about their academic work. Students upload their text-based assignments to the software, which generates quiz questions tailored to the submission with the goals of increasing student engagement in learning and assessment, deterring contract cheating and automating detection to reduce instructor workload (Sikanai, n.d.). The approach taken by Auth+ is similar to viva voce assessments (or oral examinations) that provide opportunities for students to

demonstrate or verify their knowledge and skills (Okada et al., 2015; Thomas et al., 2014). Unfortunately, viva voce may be difficult to implement in larger enrollment courses given the time needed to meet with and assess each student. Software may enable the scaling up of the viva voce assessment strategy.

Each writer uses a unique word set, makes unique word choices (Coulthard, 2004) and has a distinctive writing style (e.g., Juola, 2021) that differs to some degree from other writers. These characteristics of writing should render it easier to answer questions about one's own than another's writing. If a student did not write their assignment, their ability to differentiate their ideas from those of their peers should be impaired. In contrast, a student who crafted their assignment should have better memory for precise details (i.e., word usage) than a student who did not. Perhaps more importantly, instructors could use the information generated from authorship verification software to begin conversations with students about their learning challenges. Conversations between instructors and students can help students to reflect on their writing, writing processes and self-regulated learning strategies (Rogerson, 2017). Even if academic misconduct was not committed, student-instructor conversations can lead to supplementary support and learning opportunities. Thus, assessing students' knowledge of specific content, word choice and writing style in their assignments may be an important aspect of a multidimensional approach to promoting academic integrity in addition to detecting and deterring cheating.

Enhancing learning by quizzing

Although testing is typically viewed as a strategy to assess learning, the practice can increase students' ability to retain and recall information, thereby enhancing learning (Karpicke & Roediger, 2008; Yang et al., 2021). The knowledge of an upcoming quiz may also motivate students to complete assignments independently and review them by rereading, studying and identifying grammar, style or content issues in their academic work. If students plan to submit assignments that they did not write but understand that they will be quizzed on its content, they may still choose to study the material to pass the quiz – a learning activity they might not have engaged in otherwise and that could still enhance their learning of course content. Together, quizzing along with these associated learning behaviours may enhance long-term retention of course content. The relationship between learning and authorship quizzes may also run in the opposite direction, however, as quiz scores may be influenced by individual differences in cognitive processing that are unrelated to authorship, which would then undermine the validity of quiz results. Specifically, quiz scores may be related to the use of metacognitive self-regulation strategies and working memory, regardless of whether students are tested on their academic work or the work of another student.

Metacognitive self-regulation includes students' use of strategic behaviours to plan, monitor and regulate their learning, for example, by planning study activities and reading strategies, monitoring their comprehension and adjusting their behaviour as necessary to maximise learning (Pintrich et al., 1993). Metacognitive self-regulation strategies are associated with grade point averages, suggesting that these strategies impact learning processes relevant to performance on assessments (Kosnin, 2007; Pintrich et al., 1993). Self-regulated learning is also negatively associated with procrastination (Howell & Watson, 2007). Students lower in metacognitive self-regulation may procrastinate (e.g., last-minute preparation, less editing and studying time) and be less likely to use successful learning strategies – factors unrelated to authorship – which may impact performance on tests of authorship negatively.

If students' performance on authorship quizzes is positively associated with their metacognitive self-regulation (also associated with higher academic achievement), we would expect a positive relationship between authorship indicators and assignment grades. If this is the case, using low authorship scores as an indicator of contract cheating could lead to suspicions of misconduct by students with lower grades rather than requiring more academic support – an outcome that further disadvantages the student (Pitt et al., 2021). The validity of authorship verification software results as an indicator of authorship would be uncertain if quiz scores are highly and positively associated with students' use of successful learning and study strategies more generally.

Another individual difference factor that may influence students' performance on tests of authorship is working memory – the temporarily enhanced availability of information for processing (Cowan, 1988). Working memory is integral to cognitive functions, including reasoning (Wilhelm et al., 2013) and reading comprehension (Daneman & Merikle, 1996; Palladino et al., 2001), which are both pertinent to performance on quizzes. Individual differences in working memory are associated with the ability to encode, store and retrieve information from long-term memory (Rosen & Engle, 1997; Unsworth, 2016; Unsworth et al., 2013). Working memory capacity could affect students' performance on authorship quizzes by influencing their encoding and retrieval of assignment details to and from long-term memory during quizzing. In this case, if authorship verification scores generated by software reflect students' ability to retain, recall and process recently viewed prose, whether written by them or not, then such scores may not be a valid measure of authorship.

Present study

To assess the potential value of quizzing students on their writing for encouraging learner engagement and detecting outsourced work, we studied the implementation of Auth+ within a course at a research-intensive Canadian university. Our objective was to assess the perceived usability and learning potential of Auth+ and the validity of Auth+ quiz scores to detect and deter contract cheating. We had three main research questions:

(1) Are Auth+ scores correlated with scores on a rubric of contract cheating indicators (see London and South East Academic Integrity Network, n.d.; Rogerson, 2017; TEQSA, n.d.; UC San Diego, n.d.)?
(2) Are Auth+ scores correlated with individual differences in working memory or metacognitive self-regulation that may have an undue influence on authorship scores?
(3) Do students perceive the software as an encouragement to engage more deeply in the learning associated with their assignments and/or as a cheating deterrent?

We also asked students about their experiences using the software.

Method

Participants

In the 2022 summer term, a third-year computer science course instructor required his 51 students to submit three text-based assignments to Auth+ as part of the course requirements. Following the submission of the third assignment to Auth+, the instructor shared study recruitment information with students via email and the university's learning management system (LMS). A total of 24 (47.1%) students (M_{age} = 23.3 years, SD_{age} = 3.2 years, range: 19–30 years; gender: 87.5% men, 8.3% women and 4.2% specified another identity) consented to releasing their assignments, grades and authorship verification results to us and completed an online survey. We compensated participants with a $5 e-gift card. The course instructor was unaware of which students participated in the research study and only had access to the summary of the findings once final term grades were submitted to the university registrar. The study was approved by the Research Ethics Board at the University of Manitoba, Canada. Auth+ was inspected by our institution's Access and Privacy Office to ensure copyright and privacy prior to our obtaining it for testing.

Procedure and measures

Students completed three assignments: (a) a 2-paragraph thesis statement, (b) an annotated outline in bullet-point format and (c) an 8-page final term paper (graded out of 10). Assignments (a) and (b) were treated as practice trials to ensure students became familiar with the Auth+ interface and the types of quiz questions that it generates. After submitting their assignments to their instructor via the LMS, students uploaded the assignment to Auth+ and completed the auto-generated, time-limited quiz (~ 3 min in duration). Students interested in completing this study clicked on a link to an online consent form and survey (administered via Qualtrics) that their instructor posted to the LMS announcements page. Participating in the study was voluntary and was not part of their course requirements.

Immediately after submission of a text-based assignment, Auth+ generated six questions (in each of three categories) based on all aspects of the submitted assignment and a corpus of writing consisting of submissions from other students in the same course. *Content questions* required students to identify sentences that they wrote among sentences drawn from the broader corpus (i.e., multiple-choice questions). *Style questions* required students to identify their writing style by discriminating between sentences drawn directly from their assignments and altered versions of sentences from their assignments. *Memory questions* required students to insert the correct missing word in sentences drawn directly from their assignments (i.e., fill-in-the-blank), thereby testing their recall ability. The percentage of correctly answered questions produced three subscale scores: memory, content and style. Overall familiarity scores represented the percentage of questions answered correctly; scores falling below 75% are flagged as failures in need of investigation. The software notified students about the time limit for answering each question.

The survey included 19 Likert-type items (based partially on Venkatesh & Bala, 2008) rated on a scale from *strongly disagree* to *strongly agree* (somewhat, moderately and strongly agree ratings were combined to reflect overall agreement) and two open-ended questions to assess Auth+ user experience, and perceptions that Auth+ enhanced learning was relevant to their studies and had potential to deter contract cheating (see Table 1).

Table 1
User experience of Auth+ software: Students' frequency of agreement (%) with survey statements

Variable	Disagree	Ambivalent	Agree
Perceived usefulness			
Using the online application would improve my academic performance.	37.5	37.5	25.0
Using the online application in my studies would increase my productivity.	41.6	45.8	12.5
Using the online application would make me a better student.	33.4	33.3	33.3
Using the online application would be useful in my studies.	50.0	25.0	25.0
Perceived ease of use			
The steps for using the online application were clear and understandable.	8.3	8.3	83.3[a]
Interacting with the online application does not require a lot of my mental effort.	25.0	20.8	54.2
The online application is easy to use.	4.2	4.2	91.7[a]
I found the questions that this online application generated easy to answer.	29.2	12.5	58.3
Technology anxiety			
The online application did not scare me.	58.3	4.2	37.5
Working with the online application made me nervous.	29.2	8.3	62.4
The online application made me feel uncomfortable.	33.3	12.5	54.2
The online application made me feel uneasy.	41.7	12.5	45.8
Voluntary use			
If I had access to this online application, I would use it.	45.8	25.0	29.1
I would use this online application voluntarily.	70.8	4.2	24.9
My instructors would/should require me to use this online application in their courses.	50.0	37.5	12.5
Although it might be helpful, this online application should not be compulsory in my courses.	4.2	25.0	70.8[a]
Relevance to studies			
In my studies, usage of this online application would be important.	70.9	16.7	12.5

Variable	Disagree	Ambivalent	Agree
In my studies, usage of this online application would be relevant.	58.4	20.8	20.8
This online application would be pertinent to my various assignments.	45.8	25.0	29.1
Deterrence			
This online application will deter students from submitting an assignment that others have done for them.	25.0	12.5	62.5[a]
This online application will encourage students to complete their own assignments.	16.6	20.8	62.5[a]

[a] highlight the five most frequently agreed upon statements.

The survey also included list and prose recall tasks from the Memory Assessment Scales Battery (Williams, 1991) to assess individual differences in working memory. In the prose recall task, students had 1 minute to read a passage detailing a bank robbery and then answered questions about specific details from the prose (e.g., *How many men burst into the bank?*). For the list recall task, students were presented with a list of 12 words, one at time for 3 s each. Students then recalled the words that belonged to each of four specific categories (i.e., countries, colours, bird names, cities) separately. Prose and list recall scores were calculated by summing the number of correct responses. Practice trials preceded test trials for both tasks.

Metacognitive self-regulation was assessed using the 12-item metacognitive self-regulation subscale of the Motivated Strategies for Learning Questionnaire (Pintrich et al., 1991). Students' metacognitive self-regulation was calculated by reverse scoring two items and then averaging the scores of 12 items to create one score (Cronbach's α = .77). Higher scores reflected greater metacognitive self-regulation.

To validate the authorship scores against another metric of contract cheating, the course instructor and graders also evaluated each student's assignments for 28 indicators of contract cheating (London and South East Academic Integrity Network, n.d.; Rogerson, 2017; TEQSA, n.d.; UC San Diego, n.d.). Indicators were categorised into those related to student writing (e.g., writing and content not appropriate to discipline), citing and referencing (e.g., inappropriate or irrelevant citations) and technological indicators (e.g., irregular document properties). The instructor and graders rated their level of concern for an indicator (0 = *no concern*, 1 = *some concern* and 2 = *high concern*). Instructors and graders only evaluated an indicator if they believed that they had the ability to do so. As a result, each indicator was rated by one rater. An overall score was produced by summing the ratings of all indicators. Scores could range from 0 to 56, with lower scores indicating no concern and higher scores indicating greater concern. The course instructor shared contract cheating indicator scores, along with students' grades on the final assignment with us. Only data for consenting students were retained for analysis.

Data analysis

We originally planned to examine whether Auth+ scores on the final term paper differed across student demographics (e.g., gender, language and student status) and their academic and life stress levels, experiences within teaching and learning environments, and experiences with, awareness of and concern regarding outsourcing behaviours; however, the small size of our sample prevented meaningful analysis. Therefore, we focused our analyses of responses to survey items related to user experience, metacognitive self-regulation and working memory as these were most germane to our research questions. These responses were matched to each of the four Auth+ scores, indicators of contract cheating and final assignment grades.

Differences in Auth+ scores across question types on the final term paper were examined to further understand the role of memory. To better understand the validity of authorship scores as an indicator of outsourcing, we examined their relationship with student grades and scores provided by graders and the instructor on indicators of contract cheating. We used non-parametric tests throughout because Auth+ scores were restricted in range and both Auth+ scores and contract cheating indicator scores were not normally distributed.

Results

Authorship verification scores

Students' familiarity scores on the final term paper were restricted to four possible percentages (i.e., 50.0 [*n* = 3], 66.7 [*n* = 6], 83.3 [*n* = 9], 100 [*n* = 5]), with 9 classified as failures. Because recall is more difficult than recognition (Uner & Roediger, 2022), we suspected that students' performance on *Memory* questions would be poorest relative to the other question types that only required recognition. The omnibus test of Related-Samples Friedman's Analysis of Variance by Ranks test supported our prediction (χ^2 = 14.04, *p* < .001; see Figure 1). Pairwise comparisons using a Bonferonni correction to control the overall familywise error rate revealed that students answered fewer memory than content questions correctly (χ^2 = -.818, *p* = .02). There was no evidence that accuracy on memory and style, nor style and content questions (χ^2 = -.205, *p* = 1.0; χ^2 = .614, *p* = .13, respectively) differed.

Working memory

Familiarity scores were significantly related to prose recall scores [$\tau_b(23)$ = .39, *p* = .03], accounting for 33% of the variance (Kendall, 1970, p. 126, as cited in Walker, 2003), but this was not the case for the relationship between familiarity scores and list recall scores [$\tau_b(23)$ = .27, *p* = .11].

A total of 11 students described their views about the role of memory in quiz scores. Students stated that the Auth+ questions were more difficult to answer when the submitted assignments had more words, and when more time had passed since they wrote the assignments. Students felt that their memory for specific details (e.g., word choices) was weak and not related to their authorship or understanding of the course material and assignments. Other students mentioned that it was difficult to read, compare and differentiate large sections of text used in the multiple-choice response options within the time allotted to respond to the questions. The high level of specificity in the quiz questions and time limits to answer questions contributed to feeling anxious.
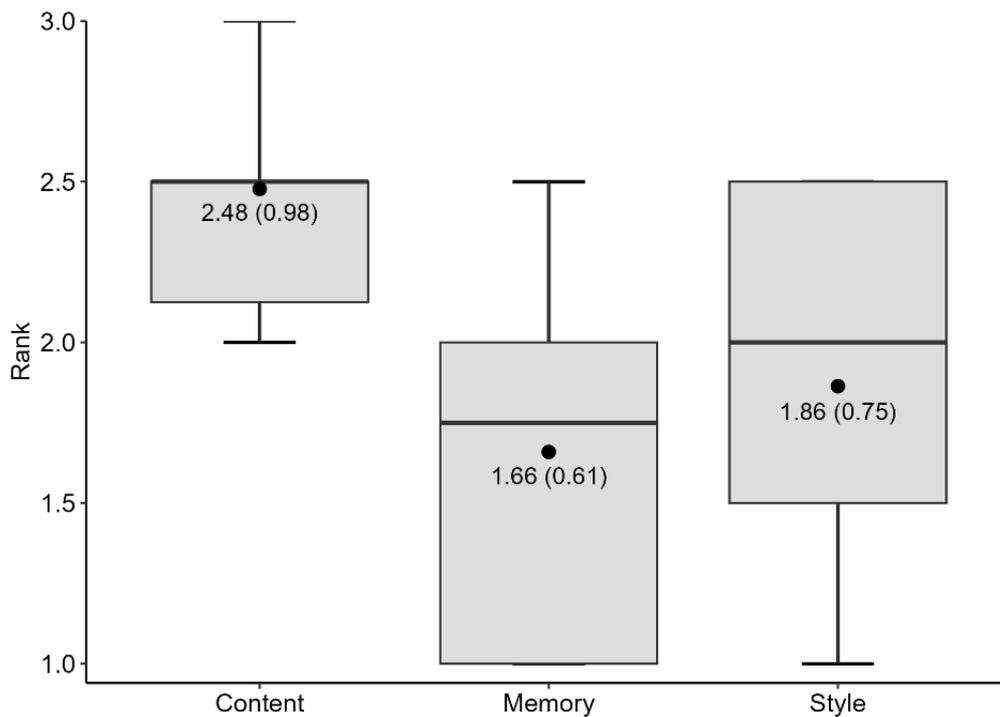

Figure 1. Mean and median ranked proportion of correct responses by question
Note. Points represent mean ranked proportions of questions answered correctly with mean unranked proportions in brackets.

Metacognitive self-regulation, grades and contract cheating indicators

Contrary to expectations, there was no evidence that familiarity scores were related to metacognitive self-regulation [$\tau_b(23)$ = .13, *p* = .44], grades on the term paper [$\tau_b(23)$ = .26, *p* = .15] or indicators of contract cheating [$\tau_b(23)$ = -.17, *p* = .35]. Thus, authorship verification scores from Auth+ did not appear to be associated with students' use of successful learning and study strategies (*M* = 4.4, *SD* = 0.9, range = 2.5 – 6), quality of assignments (*M* = 8.1, *SD* = 1.3, range = 5.5 – 10) or other indicators of outsourcing academic work (*M* = 3.3, *SD* = 3.4, range = 0 – 12). Interestingly, the number of indicators of contract cheating were negatively associated with grades attained on the final term paper [$\tau_b(24)$ = -.56, *p* < .001], suggesting that the higher the quality of students' academic work, the lower the probability that the work was outsourced.

User experience

Overall, students found Auth+ easy to use and the steps for using the application were clear and understandable. Most students agreed that the online application would deter contract cheating and encourage honest completion of assignments. Few students agreed that the application would be important for their studies or improve their academic performance. Most students also indicated that feeling nervous and uncomfortable when using Auth+. Considering students' lack of perceived usefulness and the increased anxiety they experienced using the application, it is not surprising that few students indicated that they would use the application voluntarily (see Table 1).

Discussion

Ensuring that students complete their assessments with integrity is essential. Failure to demonstrate the knowledge and skills associated with the completion of courses and programmes of study can be detrimental to students' educational pathways and their future careers. Undetected academic misconduct can also lead to a devaluing of granted credentials by employers and undermines the credibility of the educational system (Lord Ferguson et al., 2022; Yorke et al., 2020). The results of this exploratory study provide some preliminary evidence that the technology, Auth+, may be a useful component of a multifaceted approach to ensuring the validity of assessments. Two-thirds of our study participants indicated that the authorship verification software would discourage students from submitting outsourced assignments and encourage students to complete their assignments independently. Despite this promising finding, additional development of the software is required, as we found several limitations of the application.

Validity of Auth+ as a contract cheating detection tool

We found that Auth+ familiarity scores were not correlated with other contract cheating indicators described in the literature, suggesting limited validity of scores generated from the quizzing software to identify outsourced academic work. Even this finding, however, should be interpreted with caution as the contract cheating indicators (i.e., inappropriate content, errors in citing and referencing and irregular document properties) used are largely unvalidated and some may be associated with writing quality and ability rather than the probability of outsourcing (TEQSA, n.d.). Other indicators have also been studied. Research findings from an Australian study have suggested that identifying patterns of performances on proctored and unproctored assessments (i.e., differences between grades of the two assessment types) could indicate outsourced academic work (Clare et al., 2017). Again, as with other strategies (Dawson et al., 2020; Ison, 2020; Juola, 2017), this method requires comparisons across assignments for an individual and between assignment types across students.

An additional concern highlighted by our results is the influence of memory on Auth+ scores, which may impact all students, regardless of the origin of their submission. The significant positive association between working memory capacity (as measured by prose recall scores) and authorship verification scores and participant comments suggested that the use of the application for contract cheating detection

could be problematic. Participants stated that Auth+ questions were more difficult to answer when submissions were longer and more time had passed between assignment completion and submission. Students are more likely to remember the meaning of what they wrote than the details and this difference grows with a longer period between writing or studying assignments and quiz completion. Recognition memory is not simply verbatim (i.e., capturing the precise details) but is also gist-based, capturing the overall meaning (Brainerd & Reyna, 1990). Memory for precise details in prose is both worse initially and fades faster over time than does memory for the overall meaning (Sacripante et al., 2023). Instructors alerting students in advance that recalling precise words and identifying precise sentences will be expected when completing an authorship verification quiz may help. Priming people to focus on precise (surface) level details during exposure or study improves verbatim memory (Murphy & Shapiro, 1994). Yet, an authorship verification approach that includes students' ability to identify the overall meaning in their writing (e.g., thesis statement and evidence) may mitigate the impact of working memory on authorship verification scores. The next phase of the development of authorship verification software could harness the power of GenAI to analyse students' submissions to construct questions focused on meaning. Adopting any tools requiring students' submissions necessitates that institutions consider issues of copyright and privacy (which we did for our study), and instructors use their professional judgement regarding the applicability of questions generated when interpreting results obtained from authorship verification tools.

Further complicating the face validity and usability of the tool, our participants commented on the high level of unfairness inherent in the software's tendency to construct questions and answers based on quotes, citations and references within their assignments. These parts of writing are inappropriate when verifying authorship as these words "belong" to the original source and not to the students, making identification of the students' writing style and content more difficult. Difficulty answering Auth+ quiz questions with this type of content likely contributed to higher levels of anxiety. Although nervousness associated with taking tests is quite typical, anxiety before, during, and after test taking is not and interferes with learning, resulting in lower academic test scores (see von der Embse et al., 2018, for a meta-analysis). Participants also felt nervous and uncomfortable working within the Auth+ application, which is understandable when the outcome could lead to accusations of contract cheating (although in our study, this was not an outcome). Some sense of nervousness may be required for the software to be effective in one of its goals – to prevent academic misconduct.

Given the results of this exploratory study and the dearth of research on this approach to academic integrity, instructors should think carefully about the various consequences associated with implementing this type of technology, especially if it is considered as a sole source of information about authorship and cheating on assessments. Any approaches to upholding academic integrity that are considered should be scrutinised before possible implementation to ensure they are just (Moriarty & Wilson, 2022) and equitable (Steeves, 2024). More research is needed to understand the potential role that Auth+ and similar applications can play in addressing academic misconduct. The present study offers an initial investigation into the implementation of an authorship verification application in an actual university class, where the stakes could have been real for students.

Limitations and future research

Our exploratory study is not without limitations. We examined the perceptions and experiences of a small sample from a single third-year class, field and discipline, which are not necessarily generalisable to students in other years of study, fields or disciplines. Further, knowing that the study was examining experiences with software designed to detect contract cheating, students who consented to participate (vs. those who did not) may have been less likely to outsource assessments. This, combined with the limited number of quiz questions generated by the authorship verification tool, contributed to the restricted range in Auth+ scores that we observed. Future research using a larger sample that spans across fields and disciplines and increasing the number of quiz questions is needed to replicate and expand upon our findings.

Moreover, the results of the present study are correlational and provide indirect evidence regarding the validity of Auth+ in verifying authorship and detecting contract cheating, by showing the relationship between quiz scores and an individual difference factor (i.e., working memory) irrelevant to authorship, and the lack of a relationship with other indicators of academic misconduct. We have just completed a series of laboratory-based experiments where we manipulated submission of one's own or another's writing (or GenAI-produced work) to Auth+ and the ability to study the writing before submission to examine authorship verification scores across conditions (Budhiraja et al., 2024; Wong et al., 2024). Preliminary results indicate that engagement in contract cheating can be masked by studying the writing that is outsourced or AI-generated before submission as quiz scores were sufficiently high to suggest authorship (and avoid detection). Enhanced engagement with the assignment through studying may help all students succeed on a test of authorship, as studying provides time to make meaning from the content of the assignment and encode detail (e.g., constructing mental images) to enhance storage in long-term memory (Unsworth, 2016). Thus, a positive learning outcome results despite arising from a negative academic misconduct situation, which is problematic when scores are similar to those from students who wrote their assignments independently whether they studied or not. Such an outcome undermines the tools' validity for authorship verification and cheating detection.

The present study focused on contract cheating and was conducted in May and June 2022, prior to the broad public release of ChatGPT in November 2022 (OpenAI, 2022). Since the launch of ChatGPT and similar tools, the implications of GenAI for academic integrity has become a serious concern in higher education. Such concern is warranted given the prevalence of students' use of GenAI for completing assessments (i.e., 30%; Gruenhagen et al., 2024). Recent research suggests that people are not accurate in differentiating between human and GenAI writing (Kumar & Mindzak, 2024). Similarly, GenAI text detectors are not sufficiently reliable in detecting AI-generated content, are even less reliable when that output is modified and may have problematically high false positive rates (Perkins et al., 2024; Weber-Wulff et al., 2023). Reliance on GenAI text detectors may introduce bias against people whose primary language is not English (Liang et al., 2023). Given these issues, additional research should be conducted to examine the quizzing approach for authorship verification when GenAI has been used.

Lastly, we found that the majority of students reported that Auth+ did not contribute to their learning or academic performance; however, benefits may have existed that students did not perceive. Indeed, students typically do not believe that new, less familiar or unpreferred pedagogical approaches have learning value (e.g., active learning; Deslauriers et al., 2019). A deeper examination of how and under what circumstances authorship verification software that quizzes students increases engagement, even for those who wrote their assignments independently, and links to enhanced learning is warranted.

Conclusions

Our investigation of Auth+ underscores the need to verify the effectiveness of educational software aimed at detecting academic misconduct. In its current form, issues with student anxiety and a problematic relationship with working memory raise doubts about the usefulness of Auth+ to detect contract cheating. No single source of evidence should ever be relied upon to identify contract cheating, nor should educators and institutions rely solely upon software-driven strategies for promoting and upholding academic integrity. Effective technologies, however, could be one component of a comprehensive approach for promoting academic integrity and reducing temptations to engage in academic misconduct. Such an approach should endeavour to create a culture of academic integrity at the course and institutional levels and educate (i.e., how to act with integrity) and engage students, faculty and staff in its promotion (Kenny & Eaton, 2022; Morris, 2016; Stephens, 2016). A comprehensive academic integrity strategy also seeks to prevent academic misconduct through teaching excellence, deep and engaged learning and meaningful assessments (Bertram Gallant, 2008, 2017; Morris, 2016; Sotiriadou et al., 2020). Policies and accompanying procedures that emphasise student support and respond to academic misconduct as an educational opportunity to reflect and move forward, develop new skills and make better decisions round out an effective academic integrity framework (Bertram Gallant, 2017; Sopcak & Hood, 2022; Stephens, 2016). Within this framework, technology, such as Auth+, could offer another tool in the larger toolbox for promoting learning and safeguarding assessment integrity, but must continue to

be developed, tested and refined (iteratively) to be effective for making accurate judgements about students' assessments.

Author contributions

Author 1: Conceptualisation, Methodology, Investigation, Formal analysis, Writing – original draft, Writing - review and editing; **Author 2:** Supervision, Resources, Conceptualisation, Methodology, Formal Analysis, Writing – original draft, Writing – review and editing.

Acknowledgments

References

Alin, P. (2020). Detecting and prosecuting contract cheating with evidence – A " Doping Test " approach. *International Journal for Educational Integrity*, *16*, Article 7. https://doi.org/10.1007/s40979-020-00056-4

Bertram Gallant, T. (2008). Academic integrity in the twenty-first century: A teaching and learning imparative [Special issue]. *ASHE Higher Education Report*, *33*(5). https://doi.org/10.1002/aehe.3305

Bertram Gallant, T. (2017). Academic integrity as a teaching & learning issue: From theory to practice. *Theory into Practice*, *56*(2), 88–94. https://doi.org/10.1080/00405841.2017.1308173

Bertram Gallant, T., & Drinan, P. (2008). Toward a model of academic integrity institutionalization: Informing practice in postsecondary education. *Canadian Journal of Higher Education*, *38*(2), 25–43. https://www.cjhe-rces.ca/index.php/cjhe/article/view/508/557

Brainerd, C. J., & Reyna, V. F. (1990). Gist is the grist: Fuzzy-trace theory and the new intuitionism. *Developmental Review*, *10*(1), 3–47. https://doi.org/10.1016/0273-2297(90)90003-M

Bretag, T., Harper, R., Burton, M., Ellis, C., Newton, P., Rozenberg, P., Saddiqui, S., & van Haeringen, K. (2019). Contract cheating: A survey of Australian university students. *Studies in Higher Education*, *44*(11), 1837–1856. https://doi.org/10.1080/03075079.2018.1462788

Bretag, T., Harper, R., Burton, M., Ellis, C., Newton, P., van Haeringen, K., Saddiqui, S., & Rozenberg, P. (2019). Contract cheating and assessment design: Exploring the relationship. *Assessment & Evaluation in Higher Education*, *44*(5), 676–691. https://doi.org/10.1080/02602938.2018.1527892

Budhiraja, A., Stoesz, B. M., & Quesnel, M. S. (2024, May 22–24). *Studying the effectiveness and usability of an online authorship tool in verifying and detecting AI-generated* content [Conference presentation]. Joint International Conference on Ethics and Integrity in Academia, Gatineau, QC, Canada.

Clare, J., Walker, S., & Hobson, J. (2017). Can we detect contract cheating using existing assessment data? Applying crime prevention theory to an academic integrity issue. *International Journal for Educational Integrity*, *13*, Article 4. https://doi.org/10.1007/s40979-017-0015-4

Clarke, R., & Lancaster, T. (2006). Eliminating the successor to plagiarism? Identifying the usage of contract cheating sites. In F. Duggan (ed.), *Proceedings of the 2nd International Plagiarism Conference* (pp. 19–21)*. Northumbria Learning Press.

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International, 61*(2), 228–239. https://doi.org/10.1080/14703297.2023.2190148

Coulthard, M. (2004). Author identication, idiolect, and linguistic uniqueness. *Applied Linguistics*, *25*(4), 431–447. https://doi.org/10.1093/applin/25.4.431

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin*, *104*(2), 163–191. https://doi.org/10.1037/0033-2909.104.2.163

Curtis, G. J., McNeill, M., Slade, C., Tremayne, K., Harper, R., Rundle, K., & Greenaway, R. (2022). Moving beyond self-reports to estimate the prevalence of commercial contract cheating: An Australian study. *Studies in Higher Education*, *47*(9), 1844–1856. https://doi.org/10.1080/03075079.2021.1972093

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review*, *3*(4), 422–433. https://doi.org/10.3758/BF03214546

Dawson, P., & Sutherland-Smith, W. (2019). Can training improve marker accuracy at detecting contract cheating? A multi-disciplinary pre-post study. *Assessment & Evaluation in Higher Education*, *44*(5), 715–725. https://doi.org/10.1080/02602938.2018.1531109

Dawson, P., Sutherland-Smith, W., & Ricksen, M. (2020). Can software improve marker accuracy at detecting contract cheating? A pilot study of the Turnitin Authorship Investigate alpha. *Assessment & Evaluation in Higher Education*, *45*(4), 473–482. https://doi.org/10.1080/02602938.2019.1662884

Deslauriers, L., McCarty, L. S., Miller, K., Callaghan, K., & Kestin, G. (2019). Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. *Proceedings of the National Academy of Sciences*, *116*(39), 19251–19257. https://doi.org/10.1073/pnas.1821936116

Eaton, S. E., Conde, C. F., Rothschuh, S., Guglielmin, M., & Otoo, B. K. (2020). Plagiarism: A Canadian higher education case study of policy and practice gaps. *Alberta Journal of Educational Research*, *66*(4), 471–488. https://doi.org/10.11575/ajer.v66i4.69204

Ellis, C., van Haeringen, K., Harper, R., Bretag, T., Zucker, I., McBride, S., Rozenberg, P., Newton, P., & Saddiqui, S. (2020). Does authentic assessment assure academic integrity? Evidence from contract cheating data. *Higher Education Research & Development*, *39*(3), 454–469. https://doi.org/10.1080/07294360.2019.1680956

Foltynek, T., Bjelobaba, S., Glendinning, I., Khan, Z. R., Santos, R., Pavletic, P., & Kravjar, J. (2023). ENAI recommendations on the ethical use of artificial intelligence in education. *International Journal for Educational Integrity*, *19*(1), Article 12. https://doi.org/10.1007/S40979-023-00133-4

Gruenhagen, J. H., Sinclair, P. M., Carroll, J. A., Baker, P. R. A., Wilson, A., & Demant, D. (2024). The rapid rise of generative AI and its implications for academic integrity: Students' perceptions and use of chatbots for assistance with assessments. *Computers and Education: Artificial Intelligence*, *7*, Article 100273. https://doi.org/10.1016/j.caeai.2024.100273

Hamilton, M., & Wolsky, K. (2022). The barriers to faculty reporting incidences of academic misconduct at community colleges. In S. E. Eaton & J. Christensen Hughes (Eds.), *Academic integrity in Canada: Ethics and integrity in educational contexts* (Vol. 1, pp. 467–485). Springer. https://doi.org/10.1007/978-3-030-83255-1_24

Harper, R., Bretag, T., & Rundle, K. (2021). Detecting contract cheating: Examining the role of assessment type. *Higher Education Research & Development*, *40*(2), 263–278. https://doi.org/10.1080/07294360.2020.1724899

Howell, A. J., & Watson, D. C. (2007). Procrastination: Associations with achievement goal orientation and learning strategies. *Personality and Individual Differences*, *43*(1), 167–178. https://doi.org/10.1016/j.paid.2006.11.017

Ison, D. C. (2020). Detection of online contract cheating through stylometry: A pilot study. *Online Learning Journal*, *24*(2), 142–165. https://doi.org/10.24059/olj.v24i2.2096

Juola, P. (2017). Detecting contract cheating via stylometric methods. In I. Glendinning, T. Foltýnek, & J. Rybička (Eds.), *Proceedings of the 3rd International Plagiarism across Europe and Beyond Conference* (pp. 187–198). Mendel University. https://plagiarism.pefka.mendelu.cz/files/proceedings17.pdf

Juola, P. (2021). Verifying authorship for forensic purposes: A computational protocol and its validation. *Forensic Science International*, *325*, Article 110824. https://doi.org/10.1016/j.forsciint.2021.110824

Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science*, *319*(5865), 966–968. https://doi.org/10.1126/science.1152408

Kenny, N., & Eaton, S. E. (2022). Academic integrity through a SoTL lens and 4M Framework: An institutional self-study. In S. E. Eaton & J. C. Hughes (Eds.), *Academic integrity in Canada: Ethics and integrity in educational contexts* (Vol. 1, pp. 573–592). Springer. https://doi.org/10.1007/978-3-030-83255-1_30

Kosnin, A. M. (2007). Self-regulated learning and academic achievement in Malaysian undergraduates. *International Education Journal*, *8*(1), 221–228. https://openjournals.library.sydney.edu.au/IEJ/article/view/6778/7423

Kumar, R., & Mindzak, M. (2024). Who wrote this? Detecting artificial intelligence-generated text from human-written text. *Canadian Perspectives on Academic Integrity*, *7*(1), 1–9. https://doi.org/10.55016/ojs/cpai.v7i1.77675

Lancaster, T., & Clarke, R. (2016). Contract cheating: The outsourcing of assessed student work. In T. Bretag (Ed.), *Handbook of academic integrity* (pp. 639–654). Springer. https://doi.org/10.1007/978-981-287-098-8_17

Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). GPT detectors are biased against non-native English writers. *Patterns, 4*(7), Article 100779. https://doi.org/10.1016/j.patter.2023.100779

Lines, L. (2016). Ghostwriters guaranteeing grades? The quality of online ghostwriting services available to tertiary students in Australia. *Teaching in Higher Education*, *21*(8), 889–914. https://doi.org/10.1080/13562517.2016.1198759

London and South East Academic Integrity Network. (n.d.). *Contract cheating detection for markers: Checklist of strategies and features commonly found in assignments produced through contract cheating*. https://rise.articulate.com/share/dPC3F7wAQgeKahu71aUg0vBKfEUg8vsj#/lessons/6hD7HNjwZJ9vejUuxjmR5BsReX9s3au2

Lord Ferguson, S., Flostrand, A., Lam, J., & Pitt, L. (2022). Caught in a vicious cycle? Student perceptions of academic dishonesty in the business classroom. *International Journal of Management Education*, *20*(3), Article 100677. https://doi.org/10.1016/j.ijme.2022.100677

Moriarty, C., & Wilson, B. (2022). Justice and consistency in academic integrity: Philosophical and practical considerations in policy making. *Journal of College and Character*, *23*(1), 21–31. https://doi.org/10.1080/2194587x.2021.2017971

Morris, E. J. (2016). Academic integrity: A teaching and learning approach. In T. Bretag (Ed.), *Handbook of academic integrity* (pp. 1037–1053). Springer. https://doi.org/10.1007/978-981-287-098-8_11

Murphy, G. L., & Shapiro, A. M. (1994). Forgetting of verbatim information in discourse. *Memory & Cognition*, *22*(1), 85–94. https://doi.org/10.3758/BF03202764

Okada, A., Scott, P., & Mendonça, M. (2015). Effective web videoconferencing for proctoring online oral exams: A case study at scale in Brazil. *Open Praxis*, *7*(3), 227–242. https://doi.org/10.5944/openpraxis.7.3.215

OpenAI. (2022, November 30). *Introducing ChatGPT*. https://openai.com/index/chatgpt/

Palladino, P., Cornoldi, C., De Beni, R., & Pazzaglia, F. (2001). Working memory and updating processes in reading comprehension. *Memory & Cognition*, *29*(2), 344–354. https://doi.org/10.3758/BF03194929

Perkins, M., Roe, J., Vu, B. H., Postma, D., Hickerson, D., McGaughran, J., & Khuat, H. Q. (2024). Simple techniques to bypass GenAI text detectors: Implications for inclusive education. *International Journal of Educational Technology in Higher Education*, *21*, Article 53. https://doi.org/10.1186/s41239-024-00487-w

Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1991). *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ)*. University of Michigan.

Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1993). Reliability and predictive validity of the Motivated Strategies for Learning Questionnaire (MSLQ). *Educational and Psychological Measurement*, *53*(3), 801–813. https://doi.org/10.1177/0013164493053003024

Pitt, P., Dullaghan, K., & Sutherland-Smith, W. (2021). 'Mess, stress and trauma': Students' experiences of formal contract cheating processes. *Assessment & Evaluation in Higher Education*, *46*(4), 659–672. https://doi.org/10.1080/02602938.2020.1787332

Rogerson, A. M. (2017). Detecting contract cheating in essay and report submissions: Process, patterns, clues and conversations. *International Journal for Educational Integrity*, *13*, Article 10. https://doi.org/10.1007/s40979-017-0021-6

Rosen, V., & Engle, R. W. (1997). The role of working memory capacity in retrieval. *Journal of Experimental Psychology: General*, *126*(3), 211–227. https://doi.org/10.1037/0096-3445.126.3.211

Sacripante, R., Logie, R. H., Baddeley, A., & Della Sala, S. (2023). Forgetting rates of gist and peripheral episodic details in prose recall. *Memory & Cognition*, *51*, 71–86. https://doi.org/10.3758/s13421-022-01310-5

Sikanai. (n.d.). *Auth+*. Retrieved September 6, 2023, from https://www.sikanai.com/auth-plus/

Sopcak, P., & Hood, K. (2022). Building a culture of restorative practice and restorative responses to academic misconduct. In S. E. Eaton & J. C. Hughes (Eds.), *Academic integrity in Canada: Ethics and integrity in educational contexts* (Vol. 1, pp. 553–571). https://doi.org/10.1007/978-3-030-83255-1_29

Sotiriadou, P., Logan, D., Daly, A., & Guest, R. (2020). The role of authentic assessment to preserve academic integrity and promote skill development and employability. *Studies in Higher Education*, *45*(11), 2132–2148. https://doi.org/10.1080/03075079.2019.1582015

Steeves, M. (2024). Addressing equity in academic integrity. *Canadian Perspectives on Academic Integrity*, *7*(2), 1–7. https://journalhosting.ucalgary.ca/index.php/ai/article/view/77686

Stephens, J. M. (2016). Creating cultures of integrity: A multilevel intervention model for promoting academic honesty. In T. Bretag (Ed.), *Handbook of academic integrity* (pp. 995–1007). Springer. https://doi.org/10.1007/978-981-287-098-8_13

Tertiary Education Quality and Standards Agency. (n.d.). *Substantiating contract cheating: A guide for investigators*. https://www.teqsa.gov.au/sites/default/files/2022-10/substantiating-contract-cheating-guide-investigators.pdf

Thomas, J., Raynor, M., & McKinnon, M. (2014). Academic integrity and oral examination: An Arabian Gulf perspective. *Innovations in Education and Teaching International*, *51*(5), 533–543. https://doi.org/10.1080/14703297.2013.796724

UC San Diego. (n.d.). *Detecting contract cheating in narrative assignments*. https://academicintegrity.ucsd.edu/_files/take-action-files/Detecting%20Contract%20Cheating.pdf

Uner, O., & Roediger, H. L. (2022). Do recall and recognition lead to different retrieval experiences? *The American Journal of Psychology*, *135*(1), 33–43. https://doi.org/10.5406/19398298.135.1.03

Unsworth, N. (2016). Working memory capacity and recall from long-term memory: Examining the influences of encoding strategies, study time allocation, search efficiency, and monitoring abilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*, 50–61. https://doi.org/10.1037/xlm0000148

Unsworth, N., Brewer, G. A., & Spillers, G. J. (2013). Working memory capacity and retrieval from long-term memory: The role of controlled search. *Memory & Cognition*, *41*, 242–254. https://doi.org/10.3758/s13421-012-0261-x

Venkatesh, V., & Bala, H. (2008). Technology Acceptance Model 3 and a research agenda on interventions. *Decision Sciences*, *39*(2), 273–315. https://doi.org/10.1111/j.1540-5915.2008.00192.x

von der Embse, N., Jester, D., Roy, D., & Post, J. (2018). Test anxiety effects, predictors, and correlates: A 30-year meta-analytic review. *Journal of Affective Disorders*, *227*, 483–493. https://doi.org/10.1016/j.jad.2017.11.048

Walker, D. A. (2003). JMASM9: Converting Kendall's Tau for correlational or meta-analytic analyses. *Journal of Modern Applied Statistical Methods*, *2*(2), 525–530. https://doi.org/10.22237/jmasm/1067646360

Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, J., Popoola, O., Šigut, P., & Waddington, L. (2023). Testing of detection tools for AI-generated text. *International Journal for Educational Integrity*, *19*, Article 26. https://doi.org/10.1007/s40979-023-00146-z

Wilhelm, O., Hildebrandt, A., & Oberauer, K. (2013). What is working memory capacity, and how can we measure it? *Frontiers in Psychology*, *4*, 1–22. https://doi.org/10.3389/fpsyg.2013.00433

Williams, J. M. (1991). *Memory assessment scales: Professional manual*. Brainmetric Software.

Wong, K., Stoesz, B. M., & Quesnel, M. S. (2024, May 22–24). *Studying the effectiveness and usability of an online authorship tool* [Conference presentation]. Joint International Conference on Ethics and Integrity in Academia, Gatineau, QC, Canada.

Yang, C., Luo, L., Vadillo, M. A., Yu, R., & Shanks, D. R. (2021). Testing (quizzing) boosts classroom learning: A systematic and meta-analytic review. *Psychological Bulletin*, *147*(4), 399–435. https://doi.org/10.1037/bul0000309

Yorke, J., Sefcik, L., & Veeran-Colton, T. (2020). Contract cheating and blackmail: A risky business? *Studies in Higher Education*, *47*(1), 53–66. https://doi.org/10.1080/03075079.2020.1730313

Corresponding author: Matthew Quesnel, matthew.quesnel@umanitoba.ca