

Comparing ChatGPT With Experts' Responses to Scenarios that Assess Psychological Literacy

Psychology Learning & Teaching

1–16

© The Author(s) 2024



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/14757257241241592

journals.sagepub.com/home/plj**M. Anthony Machin** 

University of Southern Queensland, Australia

Tanya M. Machin

University of Southern Queensland, Australia

Natalie Gasson

Curtin University, Australia

Abstract

Progress in understanding students' development of psychological literacy is critical. However, generative AI represents an emerging threat to higher education which may dramatically impact on student learning and how this learning transfers to their practice. This research investigated whether ChatGPT responded in ways that demonstrated psychological literacy and whether it matched the responses of subject matter experts (SMEs) on a measure of psychological literacy. We tasked ChatGPT with providing responses to 13 psychology research methods scenarios as well as to rate each of the five response options that were already developed for each scenario by the research team. ChatGPT responded in ways that would typically be regarded as displaying a high level of psychological literacy. The response options which were previously rated by two groups of SMEs were then compared with ratings provided by ChatGPT. The Pearson's correlations were very high (r 's = .73 and .80, respectively), as were the Spearman's rhos (ρ 's = .81 and .82, respectively). Kendall's tau were also quite high (τ 's = .67 and .68, respectively). We conclude that ChatGPT may generate responses that match SME psychological literacy in research methods, which could also generalise across multiple domains of psychological literacy.

Keywords

Psychological literacy, ChatGPT, Situational Judgement Test

Corresponding Author:

M. Anthony Machin, University of Southern Queensland, West Street, Toowoomba, QLD 4350, Australia.

Email: tony.machin@usq.edu.au

Psychological literacy is regarded as one of the primary outcomes for all undergraduate psychology degrees (Hulme & Cranney, 2021). Recently, Cranney et al. (2022) proposed a framework for undergraduate psychology capabilities that includes both discipline-specific knowledge and skills, as well as generic capabilities that are applied in personal, professional, and community settings. Psychology is a diverse discipline, with approximately 20% of undergraduate students pursuing professional psychology careers (Hulme & Cranney, 2021). Because psychological literacy is relevant to many different pathways beyond a career as a psychologist, possessing psychological literacy can contribute to the work readiness and employable skills self-efficacy in a broad range of careers (Machin & Gasson, 2022). For students who have completed an undergraduate psychology degree, we would expect to see a moderate to advanced level of psychological literacy which supports their practice across different contexts.

Psychological literacy is not a new concept, and although the understanding of psychological literacy is still being debated, there is some convergence emerging. For instance, McGovern et al. (2010) provided nine components of psychological literacy that undergraduate students might demonstrate. However, Newell et al. (2021) concluded that there is substantial lack of agreement about what elements to include in the conceptual definition of psychological literacy, as well as whether psychological literacy has single or multiple dimensions. The same authors subsequently proposed that the key dimensions of psychological literacy included understanding, investigating, and communicating psychological knowledge (Newell et al., 2022). We accept that psychological literacy is a multidimensional construct which may therefore require a range of assessment strategies to fully capture these multiple dimensions. For the current study, we began with the assumption that psychological literacy involves the application of psychological knowledge across different contexts (e.g., personal, professional, and community as per Cranney et al., 2022).

More recently, we note that there is a disruptive influence which can potentially compromise assessment of learning in higher education, that is, generative AI tools such as ChatGPT (one example of a large language model or LLM). Ray (2023) explains that ChatGPT has the potential to revolutionise scientific research, impacting on virtually all aspects of the research process. According to OpenAI (2023), 'While less capable than humans in many real-world scenarios, GPT-4 exhibits human-level performance on various professional and academic benchmarks, including passing a simulated bar exam with a score around the top 10% of test takers' (p. 1). Therefore, it is now possible to use a generative language model to generate human-like responses to an unlimited range of scenarios. Thus ChatGPT (and other generative AI) have impacted on some of the core business of higher education and have generated debate about the appropriate role for generative AI in assessment of student learning, and this has become a major (if not the key) concern for academics across the higher education sector.

There are a number of studies which have already evaluated the performance of ChatGPT on standardised exams. For example, Choi et al. (2023) used ChatGPT to answer questions on University of Minnesota law exams and found that, while ChatGPT could pass the exams and thus be eligible for a law degree, the answers lacked deep comprehension with the materials (i.e., just at a pass level). Further, Gilson et al. (2023) discovered that ChatGPT achieved a passing score equivalent to a third-year medical student on the National Board of Medical Examiners (NBME) free 120 questions. It was able to logically justify its responses for most of the questions, although extraneous information was often included. More recently, van Dis et al. (2023) demonstrated that ChatGPT provided false information and errors in responses to summarising a systematic review, suggesting that those who use ChatGPT need to be aware of its limitations. With ChatGPT 4 more readily available (e.g., free through Bing), there is an immediate need

to evaluate the capacity of LLMs to provide the kind of responses that would exemplify psychological literacy.

Research Aims and Overview of the Two Studies

In this paper, we were interested in understanding whether ChatGPT (3.5) could provide responses that are regarded as psychologically literate. Uludag and Zhao (2023) have already demonstrated that ChatGPT can respond well to multiple choice questions used in the Graduate Record Examination (GRE) in psychology, suggesting that ChatGPT may be useful as a tutor or a study aid. Thus, we would anticipate that ChatGPT (3.5) would be able to provide basic psychological information when prompted, however, based on van Dis et al. (2023), we also were cautious as to whether this information may be incorrect or only partially true (and potentially misleading). We were also uncertain how ChatGPT (3.5) would respond to critical thinking challenges and demonstration of reflective skills, particularly in applying psychological concepts to everyday scenarios beyond multiple choice questions.

We approached the research in two different ways. That is, we were interested in how ChatGPT (3.5) would respond to scenarios used to assess psychological literacy and, subsequently, how ChatGPT's (3.5 and 4) responses compared to that of subject matter experts (SMEs) in psychology. In Study 1, we used ChatGPT (3.5) to respond to 13 open-ended research methods scenarios used to assess psychological literacy (which is the original format for the Test of Psychological Literacy (TOPL), also called the TOPL-Free Response), and in Study 2 we compared ChatGPT (3.5 and 4) to SMEs' ratings used in the development of the Revised Test of Psychological Literacy (TOPL-R). The value of comparing ChatGPT's (3.5 and 4) responses to those obtained from SMEs is that it would support the potential use of this agent (and other LLMs) in the initial training of psychology students as an enhanced, personalised tutor, which ultimately may impact on the transfer of learning to practice contexts.

Study 1

Materials

Test of Psychological Literacy (TOPL - Free Response; Roberts & Gasson, 2018). The original TOPL (Free Response) utilised 86 realistic scenarios and asked students to provide an open-ended response to the scenario along with a justification for that response. Respondents are required to answer two questions for each scenario: In response to the scenario, first, what would you say (or do/suggest/recommend), and second, what have you learned in your psychology studies that might be of relevance? In this article, we chose the research methods subscale (comprising 13 items) of the TOPL (Free Response) for both Studies 1 and 2, which has been the focus of validation efforts to date (Machin & Gasson, 2022). The 13 scenarios in this subscale are provided in Appendix 1 and focus on the application of psychological science to various research methods situations.

Method

To answer the research aim, in January 2023, we asked ChatGPT (3.5) to provide responses to 13 open-ended scenarios that demonstrate differing domains of psychological literacy. The first author inserted the actual item and the two stimulus questions (see Appendix 1) as the prompts used to task ChatGPT (3.5), and all authors discussed the corresponding output provided by ChatGPT (3.5). It

should be noted that all authors are considered experts in the field of Psychology learning and teaching and are proficient in marking student responses to the types of questions found in the TOPL (Free Response). The original TOPL (Free Response) scoring manual was also used to evaluate the responses and the scoring criteria are listed for each item in Appendix 1.

The first two scenarios that we used differed from the rest of the scenarios as there was no explicit request to identify learnings from the respondents' (in this case, ChatGPT 3.5) psychology studies that might be of relevance, however the second question still focused on providing a justification for the response. The first scenario focuses on survey responses while the second scenario focuses on designing basic research. These questions are more straightforward than other scenarios. The third scenario focused on the use of a psychic medium for a grieving friend. The fourth scenario asked about the use of alternative health treatments for a health problem. The fifth scenario asked about an organisation that wishes to use a correlation between conscientiousness and productivity as the basis for making hiring decisions. The sixth scenario referred to a negative correlation between carbohydrate consumption and cognitive performance, and how a friend has decided to remove carbohydrates from their diet. A similar situation exists for the seventh scenario, where a negative association is reported between time spent in front of screens and grades in class, and the person is asked what advice should they offer to parents? The eighth scenario describes someone who is experiencing increased stress before an exam and a friend recommends a stress reduction technique. The ninth scenario asks for advice about the mindfulness meditation diet. Our tenth scenario is a situation where a person is asked to advise a parent who has a child with autism, about a potential treatment. The eleventh scenario asks about a medical appointment with an older relative who feels overwhelmed. Scenario 12 features a sports club that wishes to recruit new members and a request to provide a recommendation about an effective recruitment strategy. Finally, scenario 13 described a relative who is new to their university and who relies on Wikipedia for all their assignment references.

Results

The results from tasking ChatGPT (3.5) with the scenario and two prompts are shown in Table 1. The 13 scenarios are listed in Appendix 1 as well as the scoring criteria developed by the authors. Our scoring (out of 2) of each response is also in Table 1.

Apart from the response for the second scenario, the ChatGPT (3.5) responses were all scored as 2 out of 2. In the second scenario, the required answer was an experimental research design that ensured tasters were unaware of the coffee brand being tasted. ChatGPT (3.5) offered four suggestions which involved customers being surveyed about their preferred brands or providing feedback after tasting samples. These responses would probably assist the manager but failed to meet the required criteria for that response.

Discussion

The range of scenarios is sufficiently broad to ensure a reliable assessment of ChatGPT's (3.5) capacity to provide responses that are psychologically literate. Apart from scenario 2 (where the response was deemed inadequate), we rated the ChatGPT (3.5) responses as receiving full marks. Many of the responses contain extremely useful advice that falls outside of the scope of our definition of psychological literacy but would still be helpful. For example, in scenario three (consulting a psychic medium), the ChatGPT (3.5) response suggests that 'It might also be helpful to consider your own beliefs and feelings about the afterlife and whether you believe

Table 1. ChatGPT Response to Each Scenario and the Score from the Scoring Guide.

Item	Response	Score (out of 2)
1	<p>(a) <i>I would recommend the following steps to optimise the response rate for the survey:</i></p> <ol style="list-style-type: none"> 1. <i>Make the survey short and to the point: Volunteers are more likely to complete a survey if it is short and only asks for information that is relevant and necessary.</i> 2. <i>Use clear and concise language: Use language that is easy to understand and avoid using technical terms or jargon that may be confusing to some respondents.</i> 3. <i>Offer incentives: Consider offering incentives, such as a gift card or a chance to win a prize, to encourage volunteers to complete the survey.</i> 4. <i>Remind volunteers to complete the survey: Send reminders to volunteers who have not yet completed the survey. These reminders can be sent via email, text message, or other means of communication.</i> 5. <i>Make it easy to complete the survey: Use a user-friendly survey tool that allows volunteers to complete the survey on a variety of devices, such as smartphones, tablets, and laptops.</i> <p>(b) <i>These recommendations are based on best practices for increasing response rates to surveys. Short, clear surveys that offer incentives and are easy to complete are more likely to be completed by volunteers, increasing the response rate for the survey. Reminding volunteers to complete the survey can also help to increase the response rate.</i></p>	2
2	<p>(a) <i>Here are some suggestions for the cafe owner to consider prior to making the decision:</i></p> <ol style="list-style-type: none"> 1. <i>Conduct a survey: Ask customers which brand of coffee beans they prefer. This will give the owner valuable insight into which brand is most popular among their customers.</i> 2. <i>Sample the different brands: Offer samples of the different brands of coffee beans to customers and ask for their feedback. This will allow the owner to get a sense of which brand customers prefer based on taste and quality.</i> 3. <i>Consider the price: The owner should also consider the price of the different brands of coffee beans and how it will impact their business. If a particular brand is significantly more expensive, it may not be financially feasible to switch to that brand.</i> 4. <i>Consult with the staff: The cafe's staff members may have valuable insights and preferences when it comes to coffee beans. The owner should consider consulting with them before making a decision.</i> <p>(b) <i>The knowledge that is helping guide this advice includes an understanding of market research techniques, such as conducting surveys and collecting customer feedback, as well as an understanding of factors that may impact a business's decision to switch to a different brand of coffee beans, such as price and staff preferences.</i></p>	0
3	<p>(a) <i>When considering whether to try and contact your deceased friend through a psychic medium, it might be helpful to first understand what exactly a psychic medium is and how they claim to be able to communicate with the deceased. You could do this by doing some research online or by talking to people who have had experience with psychic mediums. It might also be helpful to consider your own beliefs and feelings about the afterlife and whether you believe that it is possible to communicate with the deceased. Additionally, you may want to consider any potential risks or drawbacks of working with a psychic medium, such as the possibility of being scammed or being given false hope.</i></p>	2

(continued)

Table I. Continued.

Item	Response	Score (out of 2)
	<p><i>Ultimately, the decision of whether to try and contact your deceased friend through a psychic medium is a personal one and should be based on your own values, beliefs, and feelings.</i></p> <p>(b) <i>From my psychology studies, I have learned that it is common for people to grieve after the loss of a loved one and to have a desire to continue communicating with them in some way. Seeking support from friends, family, and professionals can be helpful in coping with grief. However, it is important to be aware that some practices, such as attempting to contact the deceased through a psychic medium, may not be supported by scientific evidence and could potentially be harmful if they give false hope or prevent someone from properly grieving and moving on. It is always a good idea to be cautious and critically evaluate any claims or practices that are being advertised as a means of coping with grief.</i></p>	
4	<p>(a) <i>When considering whether or not to try an alternative health treatment, it is important to gather as much information as possible about the treatment and its potential risks and benefits. This can include speaking with a healthcare professional, researching the treatment online, and seeking out the experiences of others who have tried the treatment. It is also important to consider one's own values, beliefs, and preferences when making a decision about treatment.</i></p> <p>(b) <i>In my psychology studies, I have learned that it is important to approach decisions about health and treatment with an open and critical mind, and to carefully evaluate the available evidence before making a decision. It is also important to consider the potential risks and benefits of any treatment, and to be aware of the potential for bias or misinformation when seeking out information about treatment options.</i></p>	2
5	<p>(a) <i>If I were in this situation, I would want to consider the following points:</i></p> <ul style="list-style-type: none"> • <i>It is important to consider whether the data being presented is based on a large, diverse sample that is representative of the population of potential employees.</i> • <i>It is also important to consider whether there may be other factors that could be contributing to the observed correlation between conscientiousness and productivity. For example, individuals who are more conscientious may also possess other characteristics (e.g., motivation, intelligence) that could be contributing to their productivity.</i> • <i>It is possible that the correlation between conscientiousness and productivity could be influenced by the specific tasks or job duties being performed. Some tasks may require higher levels of conscientiousness to be completed effectively, while others may not.</i> • <i>It is important to consider whether relying solely on measures of conscientiousness in hiring decisions could result in a lack of diversity in the workforce, as certain groups may score differently on these measures.</i> <p>(b) <i>In my psychology studies, I have learned that conscientiousness is one of the 'big five' personality traits, which are often used to describe and understand human personality. Conscientiousness is characterised by a tendency to be organised, reliable, and responsible. Research has generally found that individuals who are high in conscientiousness tend to be more successful in their academic and professional endeavors. However, it is important to note that personality is just one factor that can</i></p>	2

(continued)

Table I. Continued.

Item	Response	Score (out of 2)
	<i>influence an individual's productivity, and it is important to consider other factors such as skills, experience, and motivation when making hiring decisions.</i>	
6	<p data-bbox="193 382 901 409">(a) <i>If I were speaking to my friend in this situation, I would say the following:</i></p> <p data-bbox="238 409 1023 555"><i>It is important to be cautious about drawing conclusions based on a single study, especially if it has not yet been replicated by other researchers. It is possible that the observed negative correlation between carbohydrate consumption and cognitive performance could be due to other factors, or could be specific to the particular sample of individuals studied in this research.</i></p> <p data-bbox="238 555 1023 728"><i>It is also important to consider the potential drawbacks of cutting all carbs from your diet. While some research suggests that low-carb diets may be effective for weight loss and other health outcomes, it is important to ensure that you are getting sufficient nutrients from other sources. Carbohydrates are an important source of energy for the body and brain, and cutting them out completely may result in inadequate intake of essential nutrients.</i></p> <p data-bbox="238 728 1023 846"><i>It is generally recommended to follow a balanced diet that includes a variety of foods from all food groups, rather than eliminating any one group completely. It may be more beneficial to focus on choosing high-quality carbs, such as whole grains, fruits, and vegetables, rather than cutting carbs out entirely.</i></p> <p data-bbox="193 846 1023 1028">(b) <i>In my psychology studies, I have learned that nutrition is an important factor that can influence cognitive performance. Some research suggests that certain nutrients, such as omega-3 fatty acids and antioxidants, may be beneficial for cognitive function. However, it is also important to note that the relationship between diet and cognitive performance is complex, and more research is needed to understand the specific ways in which different nutrients and dietary patterns may influence cognitive function.</i></p>	2
7	<p data-bbox="193 1055 901 1082">(a) <i>If I were speaking to my friend in this situation, I would say the following:</i></p> <p data-bbox="238 1082 1023 1228"><i>It is important to consider the limitations of the research being described in the news segment. It is possible that the observed negative association between screen time and grades could be due to other factors, or could be specific to the particular sample of students studied in this research. It is also important to consider the specific context in which the screen time is occurring, as well as the content being viewed.</i></p> <p data-bbox="238 1228 1023 1374"><i>It may be more beneficial to focus on helping parents and students to use screens in a balanced, healthy way, rather than simply limiting screen time. This could involve setting limits on certain types of screen use (e.g., social media, gaming), setting aside specific times of day for screen use, and encouraging other activities (e.g., physical exercise, social interaction) that may have positive impacts on learning and development.</i></p> <p data-bbox="238 1374 1023 1465"><i>It is important to keep in mind that screens can be a valuable tool for learning and communication, and it may not be necessary or practical to eliminate screen time entirely. It is important to find a balance that works for each individual child and family.</i></p> <p data-bbox="193 1465 1023 1574">(b) <i>In my psychology studies, I have learned that the relationship between screen time and academic performance is complex and multifaceted. Some research has suggested that excessive screen time may be negatively associated with academic performance, but other research has found no such association. It is possible that the effects of screen time</i></p>	2

(continued)

Table I. Continued.

Item	Response	Score (out of 2)
	<i>on academic performance may depend on the specific content being viewed, the context in which screen time is occurring, and the individual characteristics of the student. It is important to consider these factors when examining the relationship between screen time and academic performance.</i>	
8	<p data-bbox="181 444 1005 496">(a) <i>If I were in this situation, I would consider the following points when deciding whether or not to use a stress reduction technique:</i></p> <p data-bbox="224 502 1005 760"><i>It is important to be cautious about using techniques or methods that have not been proven to be effective through scientific research. While it is possible that a technique could be helpful for one individual, it may not necessarily be effective for everyone. It is a good idea to consider the source of the information about the technique and whether it is reliable. It is generally more trustworthy to use techniques that have been recommended by mental health professionals or that have been supported by research. It is important to consider whether the technique aligns with your personal values, beliefs, and goals. If a technique does not feel comfortable or appropriate for you, it may not be effective in reducing your stress.</i></p> <p data-bbox="181 766 1005 819">(b) <i>In my psychology studies, I have learned that there are a number of evidence-based techniques that can be effective in reducing stress. These techniques include:</i></p> <ul data-bbox="224 824 624 964" style="list-style-type: none"> <i>• Deep breathing exercises</i> <i>• Progressive muscle relaxation</i> <i>• Mindfulness meditation</i> <i>• Exercise and physical activity</i> <i>• Time management and organisation skills</i> <p data-bbox="248 970 1005 1106"><i>It is generally recommended to try a variety of stress reduction techniques and to see what works best for you as an individual. It is also important to remember that stress is a normal and necessary part of life, and it is not always possible to eliminate stress completely. It is more realistic to aim for stress management, rather than stress elimination.</i></p>	2
9	<p data-bbox="181 1148 1005 1201">(a) <i>If I were considering trying the mindfulness meditation diet, I would consider the following points:</i></p> <p data-bbox="224 1206 1005 1343"><i>It is important to be cautious about following fad diets, as they may not be based on sound scientific principles and may not be sustainable or healthy in the long term. It is generally more effective to adopt a balanced, healthy eating pattern that includes a variety of foods, rather than eliminating certain food groups or following a strict, restrictive diet.</i></p> <p data-bbox="224 1348 1005 1437"><i>It is a good idea to consult with a healthcare professional or a registered dietitian before making any significant changes to your diet. They can provide guidance on safe and effective ways to reach your weight loss goals.</i></p> <p data-bbox="224 1443 1005 1523"><i>It is important to consider whether the mindfulness meditation diet aligns with your personal values, beliefs, and goals. If the diet does not feel comfortable or appropriate for you, it may not be effective in helping you to lose weight.</i></p> <p data-bbox="181 1528 1005 1581">(b) <i>In my psychology studies, I have learned that mindfulness meditation is a form of meditation that involves paying attention to one's present moment experience in a</i></p>	2

(continued)

Table I. Continued.

Item	Response	Score (out of 2)
	<p><i>non-judgemental way. Some research suggests that mindfulness meditation may be associated with a number of health benefits, including stress reduction and improved physical health. However, it is important to note that mindfulness meditation is just one aspect of a healthy lifestyle and should not be used as a sole means of weight loss. A healthy diet and regular physical activity are also important for maintaining a healthy weight.</i></p>	
10	<p>(a) <i>If I were speaking to the parent in this situation, I would say the following:</i> <i>It is important to be cautious about using treatments or interventions that have not been proven to be effective through scientific research. While it is possible that a particular treatment could be helpful for one individual, it may not necessarily be effective for everyone.</i> <i>It is a good idea to consult with a healthcare professional or a licensed mental health provider before making any decisions about treatment. They can provide guidance on safe and effective interventions for autism.</i> <i>It is important to consider the potential drawbacks and risks of any treatment, as well as the potential benefits. It is generally recommended to use interventions that have been shown to be safe and effective through research.</i></p> <p>(b) <i>In my psychology studies, I have learned that autism is a neurodevelopmental disorder that affects communication and social interaction. There is no one-size-fits-all treatment for autism, and the most appropriate intervention will depend on the individual's needs and characteristics. Some evidence-based treatments for autism include:</i></p> <ul style="list-style-type: none"> <i>• Applied behaviour analysis (ABA)</i> <i>• Early intensive behavioural intervention (EIBI)</i> <i>• Social skills training</i> <i>• Speech and language therapy</i> <p><i>It is generally recommended to use evidence-based treatments that have been shown to be safe and effective through research. It is also important to remember that individuals with autism may benefit from a combination of different interventions, tailored to their specific needs.</i></p>	2
11	<p>(a) <i>If I were in this situation, I would do the following:</i> <i>I would ask the specialist to provide more information about the potential benefits, risks, and side effects of each treatment option. It is important to have a clear understanding of the potential consequences of each treatment in order to make an informed decision.</i> <i>I would ask the specialist to explain the specific reasons why they are recommending each treatment option. This could help to clarify which option may be more appropriate for my aunt's individual situation.</i> <i>I would encourage my aunt to ask any questions she may have about the treatment options and to discuss her preferences and concerns with the specialist. It is important to involve my aunt in the decision-making process and to respect her autonomy.</i></p> <p>(b) <i>In my psychology studies, I have learned that informed consent is an important principle in medical treatment. Informed consent refers to the process of understanding and</i></p>	2

(continued)

Table I. Continued.

Item	Response	Score (out of 2)
	<p><i>agreeing to treatment, based on the availability of accurate and sufficient information. It is important to ensure that individuals have the opportunity to ask questions and to fully understand the potential risks, benefits, and alternatives to treatment before making a decision. It is also important to respect the autonomy of individuals and to allow them to make decisions about their own medical care.</i></p>	
12	<p>(a) <i>If I were in this situation, I would go about sourcing information on the two methods as follows:</i></p> <p><i>I would start by searching for research studies or articles that have examined the effectiveness of local letterbox drops and advertising on social media for recruiting new members to sports clubs.</i></p> <p><i>I would try to use reliable sources of information, such as peer-reviewed journals or reputable websites.</i></p> <p><i>I would also try to gather information from other sources, such as sports clubs that have used these methods in the past. I could ask for their experiences and insights on which method was more effective for them.</i></p> <p>(b) <i>In my psychology studies, I have learned that there are a number of factors that can influence the effectiveness of advertising and marketing campaigns. Some of these factors include:</i></p> <p><i>The target audience: It is important to consider the characteristics of the intended audience and how they may respond to different advertising methods.</i></p> <p><i>The message: The content and style of the message can influence how it is received by the audience.</i></p> <p><i>The medium: Different advertising methods may be more or less effective depending on the specific characteristics of the medium (e.g., social media, print, television).</i></p> <p><i>It is generally recommended to consider these factors and to use a combination of methods, rather than relying on a single method, in order to reach a wider audience and increase the chances of success.</i></p>	2
13	<p>(a) <i>If I were in this situation, I would advise my relative to do the following:</i></p> <p><i>Use a variety of sources when researching for an assignment. While Wikipedia can be a useful starting point, it is generally not considered to be a reliable source of information for academic work. It is important to use a mix of primary and secondary sources, such as peer-reviewed journal articles, books, and reputable websites.</i></p> <p><i>Check the credibility of the sources being used. It is important to ensure that the sources being used are reliable and accurate.</i></p> <p><i>Follow the citation guidelines for the assignment. It is important to properly cite all sources used in the assignment, using the appropriate citation style.</i></p> <p>(b) <i>In my psychology studies, I have learned that the use of credible and reliable sources is an important aspect of academic writing. It is important to use sources that have been peer-reviewed and are widely accepted within the field, in order to ensure the accuracy and credibility of the information being presented. Proper citation of sources is also important in order to give credit to the original authors and to allow readers to easily locate the sources used. Failing to properly cite sources can be considered plagiarism, which is a serious offense in academia.</i></p>	2

that it is possible to communicate with the deceased. Additionally, you may want to consider any potential risks or drawbacks of working with a psychic medium, such as the possibility of being scammed or being given false hope’.

We also examined the responses for any misleading information or ‘hallucinations’. The latter is a phenomenon that appears to occur when a LLM creates a fact out of nothing. We identified several statements that may be misleading such as ‘Some research suggests that certain nutrients, such as omega-3 fatty acids and antioxidants, may be beneficial for cognitive function’. This statement sits along with a disclaimer ‘However, it is also important to note that the relationship between diet and cognitive performance is complex, and more research is needed to understand the specific ways in which different nutrients and dietary patterns may influence cognitive function’. Therefore, we did not regard this as misleading. Another potentially misleading statement was ‘Some research suggests that mindfulness meditation may be associated with a number of health benefits, including stress reduction and improved physical health’. It also is accompanied by a disclaimer ‘However, it is important to note that mindfulness meditation is just one aspect of a healthy lifestyle and should not be used as a sole means of weight loss. A healthy diet and regular physical activity are also important for maintaining a healthy weight’.

We concluded that ChatGPT’s (3.5) responses demonstrated high levels of psychological literacy for the items focused on psychology research methods scenarios. ChatGPT (3.5) provided answers that were cautious, but clearly outlined risks and benefits based on psychological knowledge. It may be less capable in providing responses to the other subscales of the TOPL (Free Response) (e.g., psychopathology, cross-cultural psychology, and ethics) where there are more complex issues and nuanced ethical scenarios. It seems to include advice that undergraduate psychology students should be able to generate from their psychology studies. The degree of sophistication of the responses was quite high.

The second study compared ChatGPT’s (3.5 and 4) ratings of various responses to the same items from Study 1 with the SMEs’ ratings from two expert reference groups used in the TOPL-R.

Study 2

Materials

Revised Test of Psychological Literacy (TOPL-R; Machin & Gasson, 2022). The TOPL-R presents five response options for each of the 13 research methods scenarios and respondents are asked to rate the degree to which these responses reflect their own understanding and application of psychological science.

Method

Ethical approval for the collection of data for the TOPL (Free Response) and TOPL-R was provided by Curtin University Human Research Ethics Committee (HRE2017-0480-15) and the UniSQ Human Research Ethics Committee (H19REA085v5). No additional ethical approval was sought for the AI responses.

This study was a comparison of the SMEs’ ratings which were obtained as part of Machin and Gasson (2022) and Gasson et al. (2023), with those provided by ChatGPT (3.5) and ChatGPT (4). The improvement in the language model for ChatGPT (4) meant that there may be slightly different ratings and therefore we included both language models. ChatGPT (3.5) was tasked in January 2023, whilst ChatGPT (4) was tasked in November 2023. The process of collecting ratings from

SMEs is critical for weighting the responses to the scenarios in Situational Judgement Tasks (SJT). Once the SME ratings are collected, individual respondent's scores are assigned which reflect the proportion of SMEs who selected that specific rating (from 1 to 5) for that response option. There were five responses that were created for each of the 13 scenarios in the research methods subscale, and these were independently judged by two groups of SMEs ($N = 7$ and 8).

The first group of SMEs ($N = 7$) were members of a national psychology associations' interest group for psychology education. These seven respondents included four previous recipients of the award for distinguished contribution to psychology education and all were experienced psychology academics with PhDs. Their responses were collected in 2021. The second group of SMEs ($N = 8$) were either psychology academics with an APAC accredited psychology program at an Australian university or currently practising registered psychologists, representing a more diverse range of views than the first group. Their responses were collected in 2022.

Average ratings were calculated across the two groups to evaluate how similar they were with Mean of Group1 = 2.90 (SD = 1.09) and Mean of Group2 = 3.39 (SD = 1.13). To assess the degree of similarity in the two groups' ratings, the Pearson's r correlation was calculated ($r = .87, p < .001$), as well as the Spearman's rho ($\rho = .91, p < .001$). These indicate almost identical profiles in the average ratings of the SME groups. The Kendall's tau (a measure reflecting rater agreement when the metric is an ordinal scale) was also excellent ($\tau = .76, p < .001$). The SMEs' ratings were compared ChatGPT's ratings for each response across the 13 research methods scenarios.

The prompt used to task ChatGPT (3.5) and ChatGPT (4) was as follows: Consider the following scenario [insert scenario]. Please rate each Response below from 1 to 5 where: 1 = No application of psychological science and 5 = Completely appropriate application of psychological science. The SME ratings and ChatGPT's (3.5) and ChatGPT's (4) ratings for one of the TOPL-R scenarios (RM1) and its five response options is shown in Table 2. Additional Tables 3–14 are contained in Appendix 2 showing the TOPL-R stem and ratings from SMEs and ChatGPT (3.5 and 4) for research methods scenarios two to 13.

Results

The ChatGPT ratings for the responses were considerable higher than both of the SME groups' ratings with a Mean (ChatGPT3.5) = 3.97 (SD = 1.17) while Mean (ChatGPT4) = 3.80 (SD = 1.21). Group1's mean rating was 2.90 (SD = 1.09) while Group2's mean rating was 3.39 (SD = 1.13). There were many more responses that ChatGPT (both 3.5 and 4) rated as 5 (out of 5) and almost none were rated 1. We calculated the Pearson correlations with Group1 and Group2 and these were very high (r 's = .73 and .80, respectively, for ChatGPT3.5 and r 's = .77 and .84, respectively, for ChatGPT4), whilst the Spearman's rhos were also very high (ρ 's = .81 and .82, respectively, for ChatGPT3.5 and ρ 's = .81 and .85, respectively, for ChatGPT4). The Kendall's tau was also calculated for both rater groups with ChatGPT (both 3.5 and 4) to indicate the level of rater agreement where the metric is an ordinal scale. These were also quite high (τ 's = .67 and .68 for Group1 and τ 's = .69 and .71 for Group 2). These statistics were all significant at $p < .001$.

Discussion

We concluded that ChatGPT does provide a reasonable guide as to the degree of psychological literacy being required in the responses to the 13 scenarios we submitted. There may be some additional value in using the ChatGPT (4) language model as these ratings were more similar to both SME groups. The ChatGPT ratings for both version 3.5 and 4 showed a tendency towards

Table 2. TOPL-R Stem and Ratings from SMEs and ChatGPT for RMI (Machin & Gasson, 2022).

Response option	Mean SMEs Grp1	Mean SMEs Grp2	ChatGPT3.5 rating	ChatGPT4 rating
You recommend that they provide a positive incentive for participation based on research that shows this will influence behaviour.	3.71	4.13	5	4
You recommend that the survey is anonymous as this will allow respondents to answer more truthfully.	2.57	4.25	5	5
You recommend that the purpose of the survey is very clearly communicated to the volunteers, along with information around how the results will benefit them in the future (e.g., more streamlined practices, less paperwork).	4.14	4.63	5	5
You recommend that they provide a warning to participants that they should take extreme care in responding.	1.29	1.25	2	2
You recommend that the survey is made available online so make it quicker and more accessible for participants.	2.57	4.13	5	4

TOPL-RMI: a community organisation where you volunteer is planning to survey their volunteers. They ask your advice on how they can optimise the response rate for their survey.

Please rate each Response below from 1 to 5 where: 1 = No application of psychological science and 5 = Completely appropriate application of psychological science.

being more lenient than the SME's ratings which needs to be re-evaluated with further developments of the language model. We are yet to investigate whether the type of category item would impact on the quality of the ChatGPT response; that is, ChatGPT may provide poorer responses to the more complex scenarios and ethical dilemmas that are part of the larger TOPL (Free Response) and TOPL-R (with 86 scenarios across 13 categories).

ChatGPT (3.5) was first trained on information which was produced before 2021. This means that the responses provided are limited to information available before that date. However, ChatGPT (4) was trained using data available until January 2022 (see OpenAI Developer Forum (2023): What is the actual cutoff date for GPT-4?). ChatGPT and other generative AI tools are predictive models that create responses based on the prompts that are entered. While we used the prompts that students would be asked in the TOPL (Free Response) and TOPL-R, more refined prompts may produce different responses which may result in better or poorer responses from the AI. We initially used the free version of ChatGPT (i.e., 3.5) for this research, but during revisions to the paper, we repeated the analysis in Study 2 with the paid version (4). The main limitation of LLMs is that the responses are based on the strength of associations that the LLM has developed at that time, and so can vary over time.

Conclusions

In conclusion, our results demonstrate that ChatGPT's responses display moderate to high levels of psychological literacy in psychology research methods. While its responses do not yet match the rigour of a SME when rating response options, it can provide free response answers that are

sufficient to receive the maximum score. Given that ChatGPT provides promising responses which demonstrate psychological literacy in one domain, psychology educators may find ChatGPT helpful in generating assessments or providing examples for students, although caution should be applied given that ChatGPT may offer responses which contain misleading or false information (van Dis et al., 2023). Psychology educators are therefore encouraged to continue to exercise their academic judgement and expertise, and not rely solely on ChatGPT to generate assessments, particularly for important outcomes such as psychological literacy. Students who engage in utilising ChatGPT for help with assessments should also approach ChatGPT with a healthy dose of caution. While ChatGPT did provide responses that demonstrated moderate to high psychological literacy in research methods, it is too early to suggest that ChatGPT will provide similar responses in all 13 TOPL categories. Our prediction is that there may be similar levels of psychological literacy displayed across multiple domains of psychology.

When it comes to test development and the use of SMEs, ChatGPT responses are encouraging. Given that access to SMEs and other resources that were previously required to develop and test scenarios is a serious limitation for all Situational Judgement Tests, the use of ChatGPT to generate ratings in SJT is promising, especially as researchers have immediate access to the responses that can be evaluated and modified iteratively. Where there are several responses to a scenario receiving very low (indicating poor understanding) or very high ratings, these could be revised and resubmitted to ChatGPT to ensure that a wide range of response options are available. For future test development in psychological literacy, generative AI may provide a novel strategy for developing scenarios across multiple domains as well as creating examples of better and poorer responses to further provide insight into psychological literacy levels.


Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

M. Anthony Machin  <https://orcid.org/0000-0002-0967-6934>

Supplemental Material

Supplemental material for this article is available online.

References

- Choi, J. H., Hickman, K. E., Monahan, A., & Schwarcz, D. (2023, January 23). ChatGPT goes to law school. *Journal of Legal Education*. Forthcoming. <https://ssrn.com/abstract=4335905> or <http://doi.org/10.2139/ssrn.4335905>
- Cranney, J., Dunn, D. S., Hulme, J. A., Nolan, S. A., Morris, S., & Norris, K. (2022). Psychological literacy and undergraduate psychology education: An international provocation. *Frontiers in Education*, 7, 790600. <https://doi.org/10.3389/educ.2022.790600>
- Gasson, N., Knowles, A., & Machin, M. A. (2023). *Development and Validation of the Test of Psychological Literacy – Revised: Social and Developmental Psychology, and Psychopathology Subscales* [Conference

- session]. Presented at the 4th Australian Psychology Learning and Teaching (AusPLAT) Conference, Hobart, Australia.
- Gilson, A., Safranek, C. W., Huang, T., Socrates, V., Chi, L., Taylor, R. A., & Chartash, D. (2023). How does ChatGPT perform on the United States medical licensing examination? The implications of large language models for medical education and knowledge assessment. *JMIR Medical Education*, *9*(1), e45312. <https://doi.org/10.2196/45312>
- Hulme, J. A., & Cranney, J. (2021). Psychological literacy and learning for life. In J. Zumbach, D. Bernstein, S. Narciss, & P. Marsico (Eds.), *International handbook of psychology learning and teaching* (pp. 1–29). Springer International Handbooks of Education. https://doi.org/10.1007/978-3-030-26248-8_42-2
- Machin, M. A., & Gasson, N. (2022). The measurement of psychological literacy and employable skills self-efficacy in undergraduate psychology students. *Scholarship of Teaching and Learning in Psychology*. Advance online publication. <https://doi.org/10.1037/stl0000336>
- McGovern, T. V., Corey, L., Cranney, J., Dixon, W. E. Jr, Holmes, J. D., Kuebli, J. E., Ritchey, K.A., Smith, R. A., & Walker, S. J. (2010). Psychologically literate citizens. In D. F. Halpern (Ed.), *Undergraduate education in psychology: A blueprint for the future of the discipline* (pp. 9–27). American Psychological Association. <https://doi.org/10.1037/12063-001>
- Newell, S. J., Chur-Hansen, A., & Strelan, P. (2021). A construct validity analysis of the concept of psychological literacy. *Australian Journal of Psychology*, *73*(4), 462–474. <https://doi.org/10.1080/00049530.2021.1922069>
- Newell, S. J., Chur-Hansen, A., & Strelan, P. (2022). A revised definition of psychological literacy: Multiple stakeholder perspectives. *Scholarship of Teaching and Learning in Psychology*. Advance online publication. <https://doi.org/10.1037/stl0000326>
- OpenAI. (2023). *GPT-4 Technical report*. Retrieved May 9, 2023, from <https://doi.org/10.48550/arXiv.2303.08774>
- OpenAI Developer Forum. (2023). *What is the actual cutoff date for GPT-4?* Retrieved December 18, 2023, from <https://community.openai.com/t/what-is-the-actual-cutoff-date-for-gpt-4/394750>
- Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations, and future scope. *Internet of Things and Cyber-Physical Systems*, *3*, 121–154. <https://doi.org/10.1016/j.iotcps.2023.04.003>
- Roberts, L. D., & Gasson, N. (2018). Facilitating psychological literacy: The importance of conceptualisation and measurement. In G. J. Rich et al. (Eds.) *Teaching psychology around the world (Volume 4)*. Cambridge Scholars Publishing.
- Uludag, K., & Zhao, M. (2023, April 11). *Can ChatGPT answer GRE psychology questions?* <https://doi.org/10.2139/ssrn.4416365>
- Van Dis, E. A., Bollen, J., Zuidema, W., van Rooij, R., & Bockting, C. L. (2023). ChatGPT: Five priorities for research. *Nature*, *614*(7947), 224–226. <https://doi.org/10.1038/d41586-023-00288-7>

Author Biographies

Emeritus Professor Tony Machin is a psychological scientist and organisational researcher who was in academia for three decades at the University of Southern Queensland, Toowoomba, Australia. He received the APS Distinguished Contribution to Psychology Education Award in 2021. This award recognised his significant contribution to psychology education in Australia which includes several key roles that involves developing psychology education programs, assisting with accreditation and quality assurance of psychology education programs, and advancing the status of psychology education in Australia. His current research interests include understanding the way that students balance their work and study, and measurement of psychological literacy.

Associate Professor Tanya Machin is the Associate Dean (Learning & Teaching) for Academic Affairs at the University of Southern Queensland, as well as an Associate Professor in Psychology and Counselling. She is the Chair of the Australian Psychological Society's (APS) Psychology

Education Interest Group, and a committee member of the APS Division of Psychology Research, Education and Training. Her research focuses primarily on social media and technology across the lifespan, but she is also interested in human ethics and policy. In 2020, she was the recipient of an APS Early Career Teaching award.

Professor Natalie Gasson from Curtin University (Australia) is a registered psychologist, and Fellow of the Curtin Academy (Executive). She is the Psychology Level Two Coordinator and teaches both Honours Dissertation units. Her other teaching areas include adulthood/ageing, and program evaluation. She is currently involved in a program of iSoLT via an ongoing research project that is developing a reliable and valid measure of psychological literacy. Her other research areas include Parkinson's (investigating cognitive, psychological and motor heterogeneity in Parkinson's), stigma, and program evaluation.