

Not just a pretty picture: Scientific fact visualisation styles, preferences, confidence and recall

Information Visualization
2021, Vol. 20(2-3) 138–150
© The Author(s) 2021
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/14738716211027587
journals.sagepub.com/home/ivi


Erin I Walsh¹ , Ginny M Sargent¹ and Will J Grant²

Abstract

Visuals are often used to enhance learning of scientific information. The recent emergence and popularity of comic-style instruction books for adults, such as the ‘manga guide to ...’, shows the possibility of comic style visualisations for the communication of science with adults. This study investigates whether the addition and style of visual accompaniment of scientific information, as used in comic books, influences immediate and short-term fact recall in an adult audience. Participants ($n = 310$ aged 18–79, 52% identified as female) were presented 20 general science facts in one of five styles: (1) text alone, (2) photo with text caption, (3) cartoon with text caption, (4) photo with explanatory agent and a speech bubble, (5) cartoon with explanatory agent and a speech bubble. Immediate recall, and confidence in that recall, was tested following a brief distractor. Participants indicated their preferred presentation style, and short-term recall was tested by a final quiz of all 20 facts. Overall, the most preferred presentation style was cartoon with explanatory agent and text in a speech bubble (26% preferred). There was no single most effective presentation style; there was no significant difference in immediate recall, short term recall or confidence in answers depending on whether the fact was presented as text, photo or cartoon, or the presence or absence of an explanatory agent. However, immediate recall was significantly better when preference was met ($p < 0.02$). We found that the style of visual accompaniment of scientific information in accordance with the ‘manga guide to...’ format influenced immediate, but not short-term, fact recall in an adult audience when written English literacy, scientific literacy and non-verbal intelligence were taken into account. Short term recall of scientific facts may best be served by presenting facts in multiple styles, or enquiring about and then meeting participant preference for visual accompaniment.

Keywords

2D, combinations, degree of interest, design recommendations, image, infovis, knowledge tasks

Introduction

There is a long history of using graphs, photographs and illustrations to communicate scientific information, and there is similarly a long history of assessing their impact.^{1,2} In scientific communication, these visual formats can be used to either visually represent accompanying textual data, or visualise information in a new way that reinforces the text.^{3,4} Extensive research attention has been paid to the ideal characteristics and application of visuals to enhance scientific education for children.^{1,5–8} However, the recent

¹PHXchange (Population Health Exchange), Research School of Population Health, College of Health & Medicine, Australian National University, Acton, ACT, Australia

²Australian National Centre for the Public Awareness of Science, College of Science, Australian National University, Acton, ACT, Australia

Corresponding author:

Erin Walsh, PHXchange (Population Health Exchange), Research School of Population Health, College of Health & Medicine, Australian National University, 54 Mills Road, Acton, ACT 2601, Australia.

Email: erin.walsh@anu.edu.au

emergence and popularity of comic-style instruction books such as ‘The manga guide to ...’ shows the possibility of comic-style visualisations for the communication of science with adults. It remains to be established if the illustration visual style favoured in the ‘manga guide to...’ format is more, less or equally successful in promoting learning in adults. The degree to which visual elements common to this format, such as information being presented by speech bubbles and character interaction (rather than as figure captions), has not so far been explored. This study fills this gap. This study sought to clarify whether the use of comic-style visuals alongside and integrated with textual scientific facts had an effect on immediate and short-term recall in adults, and whether the nature of the relationship between visuals and text influenced recall. This study has implications for how visuals are used in educational materials to support scientific learning in adults.

Literature review

Highly visual educational books for adults have been produced in the past decade, such as the ‘manga guide to ...[statistics/calculus/physics etc]’, which consists of textbook information presented in comic book format (often with a Japanese ‘manga’ aesthetic), intended as a supplement to education at a tertiary level (Google ngram results for ‘manga guide’, accessed 10/08/2020 https://books.google.com/ngrams/graph?content=manga+guide&year_end=2019&year_start=1800&smoothing=3&corpus=26&direct_url=t1%3B%2Cmanga%20guide%3B%2Cc0#t1%3B%2Cmanga%20guide%3B%2Cc0). Research on the utility of this format has focussed on how humour, narrative and visual representation together affect learning.⁶ A regularly used characteristic of the comic book format is the inclusion of an ‘explanatory agent’, defined here as a human or anthropomorphic animal or object that communicates to other characters or the reader via speech or thought bubbles. It is ambiguous whether the inclusion of an explanatory agent in visuals, accompanied by shifting explanatory text from caption to speech bubbles, impacts information recall in adults. Related evidence suggests an explanatory agent can enhance engagement (hence encoding) of information. For example, students have a preference for the lecturer being visible in online or recorded lectures.⁹ Furthermore, anthropomorphising to introduce topic-specific explanatory agents has been a key tool in engagement with non-human topics such as robotics.¹⁰ Another characteristic of the comic book format is the use of illustrations to depict both the substantive content and explanatory agent. This is at odds with the prevalence of

photographs, rather than cartoons, in formal scientific textbooks.¹¹

Both photographs and illustrations can equally be used to decorate, motivate or visually represent salient elements of a concept.^{10,12} It is typically assumed that photographs can provide useful reference for concepts that are physically observable (e.g. the pattern of wing motions in a bird’s flight can be directly viewed), providing ecological validity by depicting the physical reality of an object (albeit a depiction curated by the photographer, as discussed by Berger.¹³ Conversely, illustration is particularly useful in depicting concepts that cannot be directly experienced (e.g. atomic relationships too small to see),⁵ and present only the elements of the visual scene that are relevant.¹⁴ Photos and illustrations can be thought of on a spectrum from realism to more abstract iconography.¹⁵ If ‘style’ is defined as in Moere et al.¹⁶ to refer to an abstract concept that relates to how a visualisation is recognised and may be categorised, photographs and illustrations of varying complexity can be considered different ‘styles’ of visual.

Learning is not a unitary concept. It consists of acquiring new knowledge, understanding, skills, behaviours, attitudes, preferences or values,¹⁷ and in an educational setting involves focus on attention, engagement, knowledge retention and recall. Recall is defined here as ‘the mental process of retrieval of information from the past’.¹⁸ This is distinct from comprehension, which is the process of meaning extraction.¹⁹ Though recall and comprehension share several mutually reinforcing processes,²⁰ this manuscript focusses primarily on recall to reflect the necessity of recall to demonstrate the occurrence of learning in education and experimental settings. In its simplest form, a test of recall involves presentation of novel information of some kind (encoding), some passage of time (storage) and then retrieval. A classic real-world example are the exams conducted in schools, designed to test the recall of content taught in class, an area that has received much attention in the education literature (e.g. see¹²).

Recall is related to, but distinguishable from working memory.²¹ Working memory serves as a highly short-term ‘buffer’ of information encoding on the order of seconds,²² primarily allowing the manipulation of information. Working memory and recall share similar cognitive processes (e.g. central executive function)²³ and neural substrates (e.g. prefrontal activity).^{24,25} However, working memory consists of multiple non-overlapping components specific to different information modalities (phonological loop, episodic buffer and visuo-spatial sketchpad),²⁶ and relies more heavily on the dorsolateral prefrontal cortex than short-term recall.²⁷ Compared with recall, working

memory is therefore more amenable to disruption from attending to other information immediately following presentation of a stimulus,²⁸ and the alignment between the form of presented information and the distracting stimulus is more pertinent to performance.²⁹

There is substantial evidence that the nature, reliability and cognitive process of memory changes with the amount of time between encoding and recall.²¹ Similarly, although all recall relies somewhat on attention and engagement and can be facilitated by comprehension,¹⁹ there is differential performance as the time between coding and recall is extended. Here, we use ‘immediate recall’ to refer to a minimal passage of time (< 1 min) between encoding and recall, and ‘short-term recall’ to refer to recall between 1 and 30 min from encoding. ‘Long term’ recall on the order of months and years is widely considered, particularly in the neurodevelopmental³⁰ and educational³¹ literature, but is beyond the scope of this manuscript.

Immediate and short-term recall can be distinguished on the basis of differential underlying cognitive processes, manifest in differential performance relating to how information is presented, and in particular contextual cues.³² Immediate recall is more closely tied with cues specific to the information presented, such as the exact wording of text accompanying an image.³³ Conversely, short-term recall relies more on an individuals’ experience relative to the information presented, as it allows the use of ‘convergent’ retrieval where cues from one memory are used to assist recall from another,³³ similar to connectionist architecture, spreading activation and mapping processes in comprehension as described in McNamara et al.¹⁹ This can improve recall, or damage it if information is paired with an incorrect or unrelated memory.³⁴ Anglin and Stevens³⁵ experimentally manipulated the use of images accompanying text in teaching university students scientific facts and procedures, and concluded that the use of images improved immediate recall (test immediately after materials were studied), but not delayed recall (test at 28 days).

This highlights the importance of considering individual characteristics when exploring learning and recall.^{1,36} Broadly, four key individual characteristics salient to recall pertinent to topics such as scientific facts are reading proficiency (general and scientific), non-verbal intelligence, prior knowledge and preference. Where information is presented in text, reading proficiency in that language – here, English – is a prerequisite to access the information. In particular, learning scientific information and reasoning requires proficiency beyond general use and interpersonal communication. Conceptualised by Cummins³⁷, ‘cognitive-academic language proficiency’ focusses on the

ability to read technical materials such as textbooks, and, particularly important for the testing of recall, the ability to produce appropriate responses to written test questions. At least in children, cognitive academic language proficiency in particular is associated with scientific learning and consequent fact recall.⁷ This should not be conflated with intellect or learning capacity, as one can score poorly on written tests but nonetheless be highly intelligent.³⁸ Nonverbal intelligence, measured by tools such as Raven’s Progressive matrices,³⁹ is therefore a useful complimentary measure to English language proficiency to account for variability across participants in learning capacity. Prior knowledge is likely to be linked with cognitive-academic language proficiency and nonverbal intelligence, which both can pose limitations to the attainment of a knowledge base. For example, the presentation of information in the context of a comic book provided recall benefits specifically in poorly performing students in grade 10 (median age of 15).⁶

Cognitive fit theory proposes that individual performance in engagement, recall and comprehension may differ according to the ‘fit’ between how information is conveyed (e.g. text, visuals, graphs or combinations thereof) and their individual cognitive characteristics, as well as the nature of the information conveyed or task undertaken.⁴⁰ If the nature of the information and/or task is held constant, the likelihood of ‘fit’ can somewhat be inferred from both performance, and an individual’s stated preferences.³⁶ This is similar to the learning styles hypothesis, which posits different individuals learn in different ways, and that learning is facilitated when information delivery matches that individual’s learning style (e.g. visual learner learns most effectively when information is presented).⁴¹

Building from the design maxim ‘attractive things work better’,⁴² the benefits of accompanying text with visualisation is particularly salient in scientific textbooks.^{43,44} Relevant imagery is generally beneficial for recall.^{41,45} The chief explanation for this is dual coding theory, which postulates that cognition (a subset of which is learning) is characterised by a word-focussed symbolic/verbal system, and an image-focussed sensimeter/visual system. This theory holds that information presented via both pathways (e.g. text accompanied with a visual) will be better encoded and recalled than information presented by one pathway.¹² Other theories suggest different mechanisms for improved recall performance when information is presented via text (and tested via written means). The first is learning-test congruence, which suggests test performance is highest when the modality in which information is learned matches the modality in which it is tested.⁴⁶ The second is that people tend to preferentially attend to text: eye tracking studies of text in

the context of comics for communication of scientific facts has indicated significantly longer dwell time on text than visual elements.⁴⁷ While the potential utility of relevant imagery is clear, it remains unclear what constitutes ‘relevant’ imagery, because recall is also linked with contextual cues not necessarily linked with substantive informational content, such as narrative⁴⁸ and humour.⁴⁹

It is unclear why ‘relevant’ imagery is apparently beneficial to recall of scientific information,⁴¹ yet comic book formats that include an explanatory agent and cartoon ‘style’ are particularly popular. It might be that the perceived benefits of that irrelevant imagery, (visual content that is not semantically related to informational content) are actually due to style. Visuals that include an explanatory agent are typically simple illustrations or cartoons, which have the benefit of simplified visual scenes containing only relevant information, while it is rare to encounter such an explanatory agent in science textbooks. This could be disambiguated by comparing preference and recall performance when an explanatory agent is included in both illustrations and photographs, or removed altogether. Another possibility is that the explanatory agent’s role is purely to increase the reader’s enjoyment, which in turn may facilitate encoding and recall independently of informational content.⁵⁰ If preference maps to cognitive fit,^{36,40} adults who generally prefer visuals including an explanatory agent should exhibit superior recall compared with when the agent is not present. Simpler styles of illustrations, such as cartoons, bootstrap recall on enhancing engagement, motivation and ‘fun’ even in absence of humour.⁵¹ Indeed, gaze tracking studies have found that simpler illustrations promote better learning and recall.⁵²

Purpose of current study

This study investigated whether the addition and style of visual accompaniment to textual information influenced immediate and short-term recall of scientific facts in adults, when controlling for cognitive-academic language proficiency, nonverbal intelligence and general scientific literacy. In particular, we conducted an experiment on the influence of visual elements common to the ‘manga guide to...’ format by testing (1) whether there are any differences in recall accuracy if a visual is presented alongside text, (2) whether a photographic or cartoon style visual influences recall and (3) whether the presence or absence of an explanatory agent (person) has any effect on recall accuracy. The scope of this study was to replicate the ‘manga guide to...’ format, rather than test the informational or semantic content of visuals accompanying information. We focussed on recall

because it is a fundamental, readily testable aspect of learning. The time between encoding and recall can impact the degree to which cues adjacent to information may impact performance.²¹ The role of visuals as a cue has been well established when contrasting short and long-term recall,³³ but it is unclear whether immediate and short-term recall performance may differ on the basis of how visuals are designed.

Method

This section describes how the study was designed and executed. We begin by describing how the pilot was used to develop, evaluate and refine the experimental protocol. After outlining the results from repeated pilot tests, we outline the methods of the main study, including development of the materials, the participant cohort, stimulus set, measures used and study procedure.

Pilot study: Choice of questions about scientific facts

This research started with a study to pilot and refine the questions that would form the basis of testing scientific fact recall in the main study. A series of scientific facts was first compiled and a series of questions about them developed. We then conducted two rounds of pilot testing to calibrate the difficulty of these questions, aiming to identify questions about scientific facts that were not general knowledge. Other aspects of the main study protocol were not pilot tested.

The number of questions required for the main study was chosen to balance the needs of statistical testing against the burden on participants. A total of 20 scientific facts gives each participant four exposures to each of the five alternative presentation modes. In order to serve their purpose as core stimulus for immediate and short-term recall of scientific information, the majority of these questions needed to be novel to the majority of participants. To achieve this, we developed new questions analogous to the PEW research centre Science and Technology quiz (henceforth PEWST), presented in a modified CLOZE procedure. As in Porter,⁵³ this is a response format where an incomplete sentence was accompanied by five possible completion options. An example is: *The _____ is the formal name for a flower’s stalk* (with the possible options of *peduncle*, *filament*, *xylem*, *style* and *propposite*). Topic area was also considered, aiming for questions from a variety of physical and natural sciences. A result of 20% correct, or lower, indicates that a scientific fact is not general knowledge, that is, less than the one in five chance of randomly guessing the correct

answer. Questions were suitable for the main study if the general knowledge was low – allowing greater scope for measuring scientific fact recall in the main study. However, we aimed to retain some variability in question difficulty, including a mix of ‘easier’ and ‘harder’ questions to avoid poor data quality that can eventuate when participants become frustrated (e.g. see⁵⁴).

These pilot tests were undertaken via online self-report questionnaire on the Qualtrics platform by adult participants from Australia, the United States of America, the United Kingdom, New Zealand and Canada who were recruited via online crowdsourcing website, Crowdfunder. Participants disclosed age, gender, first language, most used language in the past year, country of current residence and highest level of education. While data was recorded for all participants, only those who endorsed the item ‘I took this survey seriously, and answered every question to the best of my ability’ and self-reported either first or most used language as English were included in the analyses.

Participants were presented with all facts in a single block, in randomised order. The order of possible CLOZE responses was also randomised. Different individuals completed each round of pilot testing and the main study, to ensure that the first presentation of the CLOZE procedure was the first time participants had encountered the presented facts.

Pilot 01 results. Thirty-eight respondents completed the first questionnaire and data from 11 was disregarded on the basis of poor survey engagement or their reporting of not using English as a first language. The included 27 participants were aged 18–57 years (mean 32 years), 70% were female, and 48% had a university education. Candidate facts and their performance in pilot testing are exhaustively listed in Supplemental Table S2. Briefly, the average percentage correct was 64%, indicating most were too easy or familiar.

Pilot 02 results. As a result of the first test of the scientific facts, all statements were reviewed and either entirely replaced, or updated to difficult variants (e.g. *The part of the flower that makes pollen is the **stamen**, located in the middle of the flower* became *The **anther** is the part of the stamen where pollen is produced.*), resulting in a list of 35 candidate questions for the second round of pilot testing. Ninety-two respondents completed the questionnaire, data from 34 was excluded on the basis of poor survey engagement or not using English as a first language. The included 58 participants were aged 18–73 years (mean 46 years), 59% were female and 50% had a university education.

The overall average percentage correct remained high (34%). A selection of 20 scientific facts was made for the final study (Supplemental Table S2), for which the average percentage correct was 25%.

Main study

Development of materials. Materials to present each of the 20 piloted scientific facts in five style variants were prepared (e.g. Figure 1): text only, text caption with cartoon, text caption with photograph, cartoon with explanatory agent, and photograph with explanatory agent. The photographs were selected from open source repositories such as Wikimedia Commons, on the basis of key words included in the scientific facts (e.g. ‘clouds’). These images were chosen to be relevant to the scientific fact, but not to visually communicate the information independently from the text. Photographs of the person used as the explanatory agent were taken by author EW for the purposes of this study. Four variants reflecting common poses in ‘manga guide to...’ comics (eye contact with the viewer, cropped at the waist, one finger pointing at the background stimulus) were variably used for the purposes of visual variety. The explanatory agent was superimposed onto the photographs and a text bubble including the scientific fact was added. Cartoon versions of each were created by author EW. To perceptually match the photograph and cartoon as much as possible, EW directly traced the photographs for lines and used the colour picker tool for colour fills. For the full list of facts and their variants, see Supplemental Table S3.

Participants

Adult participants from Australia, the United States of America, the United Kingdom, New Zealand and Canada were recruited via online crowdsourcing website, Crowdfunder, to complete an online self-report questionnaire via the platform Qualtrics, in return for a small monetary incentive. The ethical aspects of this research were approved by the Australian National University human research ethics committee (protocol number 2014/553). While data was recorded for all respondents, only data that met the following inclusion criterion were retained for analysis: (a) genuine engagement, satisfied by completion time of 5 min or more, and endorsing the item ‘I took this survey seriously, and answered every question to the best of my ability’; and (b) English Language proficiency, identified by self-reported residence in Australia, the United States of America, the United Kingdom, New Zealand or Canada and self-reporting English as either a first

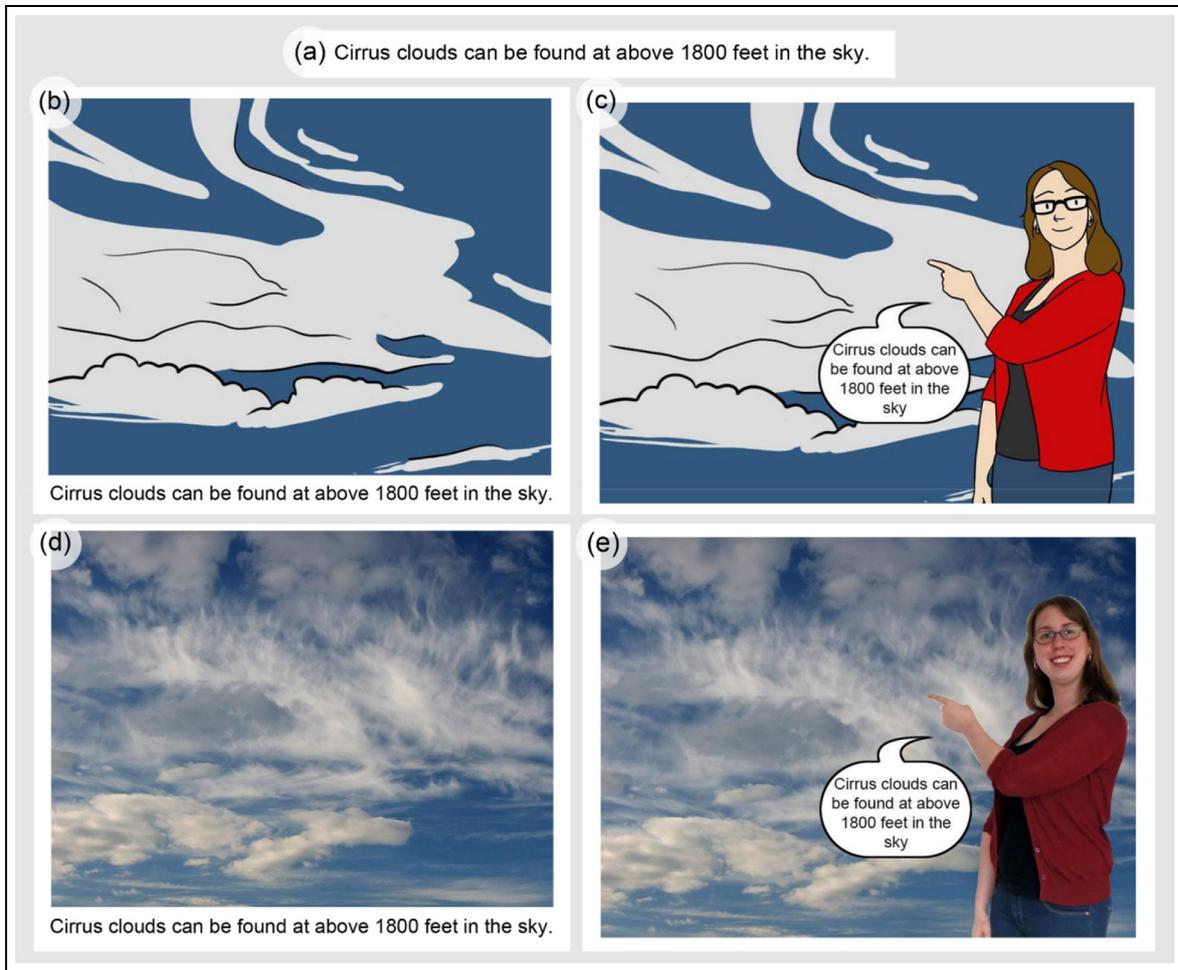


Figure 1. Example of the five style variants used to present each scientific fact: (a) text alone, (b) text accompanying cartoon, (c) cartoon with explanatory agent, (d) text accompanying photograph, and (e) photograph with explanatory agent.

Note. See Supplemental Table S3 for all variants of all questions.

language, or most spoken language. Of an original 385 respondents, data from 65 were excluded on the basis of criterion (a), and none on the basis of criterion (b), resulting in a final sample size of $n = 310$.

Measures

Demographic measures. Participants reported their age, gender, first language, the language they had used most in the previous year, their country of current residence and the highest level of education they had attained.

Covariates. A brief battery of items were included to allow insight into participant's individual capacity to understand written text (English language proficiency), prior scientific knowledge (PEWST), and

nonverbal capacity for induction and learning (Raven's progressive matrices).

English language proficiency CLOZE procedures correlate strongly with norm-referenced English language tests and language proficiency interviews.⁵⁵ Accordingly, English language proficiency was measured via a modified CLOZE procedure where participants were invited to complete a simple paragraph by selecting a correct word from a drop-down menu (Supplemental Table S1). This was scored between zero (no words correct) and five (all words correct). Science literacy was quantified with the PEW Research Centre Science and Technology quiz, which consists of 13 multiple choice questions about a wide range of science topics under the groupings of 'science in the news/daily life' (scored 0–5) and 'Textbook science' (scored 0–8), where higher scores indicate better scientific literacy.⁵⁶ Nonverbal intelligence was

quantified by a subset of six images from Raven's progressive matrices.³⁹ This is a problem-solving, abstract reasoning task where participants are presented a short sequence of images that change in a systematic way, and are then asked to select one of six possible images that would logically complete the sequence. This has a possible score of zero (implying lower nonverbal intelligence) to six (implying higher nonverbal intelligence).

Participant confidence. Confidence in responses was measured via a visual analogue scale slider, where 0 was 'Not at all sure (just a guess)' and 100 was 'Completely sure (know it is correct)'.

Participant preference. All participants responded to this question: 'You've probably noticed that we have been telling you about the scientific facts in different ways. Please select the option you preferred'.

Participant engagement. All participants responded to this question: 'To help us analyse your results, we need to know how seriously you took this survey. Please be honest – your answer here will not affect any incentive you will receive for completing this survey'.

Procedure

Participants completed all measures via Qualtrics, an online survey platform, following the procedure outlined in Table 1. Note that, rather than the more traditional approach of asking demographic and covariate questions in a single block, they were mixed and split into 20 segments of approximately equal length (e.g. age + two PEWST questions; years of education and one of the Raven's progressive matrix tasks, etc.). These were interleaved between fact presentation and immediate recall. This was done to prevent participants engaging in verbatim mental rehearsal (the temporary storage and manipulation of information), as doing so would result in data more indicative of working memory capacity than immediate recall.²⁸

Data inclusion

Questions that individuals correctly answered pre-test were excluded from subsequent analysis for that fact, for that participant. Chi square (χ^2) tests were used to check whether these exclusions resulted in an unbalanced use of the five presentation styles across participants.

Statistical analysis

A series of linear models were fit to investigate whether the method of scientific fact presentation were associated with immediate recall, short-term recall, and confidence in immediate recall. Mixed effects (also known as hierarchical) linear regression was applied to properly account for between-fact variability, such that all predictors and responses (level 1) were nested by fact (level 2). All models controlled for PEWST, cognitive-academic English proficiency scores, and Raven's progressive matrix tasks. Short-term recall additionally controlled for whether immediate recall was correct or incorrect. Alpha was set at 0.05, with Bonferroni correction for multiple comparisons reducing this to $\alpha = 0.02$, hence 98% confidence intervals.

Mixed effects models were fit using the *lme4* package,⁵⁷ with *glmer* function and binomial link for fact recall accuracy (correct/incorrect) and *lmer* function for confidence (continuous variable). Model assumption testing is described and reported in Supplemental Table S4, and Figure S1.

Sensitivity analysis. Three sets of sensitivity analyses were conducted: basic models with the key predictor only (to allow the greatest possible predictive power), models with Raven's progressive matrices and English proficiency scores treated as ordinal variables (due to their restricted range, see⁵⁸), and models where individuals with cognitive-academic English language proficiency scores < 5 were excluded (to further remove bias from poor comprehension).

Results

Participant characteristics

Participants were aged between 18 and 79, and 52% ($n = 161$) identified as female. The majority (43%, $n = 133$) currently resided in the United States of America, followed by the United Kingdom (30%, $n = 93$), Canada (22%, $n = 68$) and the remainder residing in Australia or New Zealand. All participants had completed at least a high school education, and over half (63%, $n = 195$) had completed one or more years at university. They demonstrated mixed scientific literacy as indicated by the PEWST ($M = 10$, $SD = 3$, range 0–13). Similarly, Raven's progressive matrices indicated a range of nonverbal intelligence ($M = 3$, $SD = 2$, range 0–6). The majority (79%, $n = 244$) of participants scored perfectly (5/5) in the basic English language proficiency measure. All participants were included in the main analysis, but sensitivity analyses were conducted where only those with a perfect 5/5 score were included.

Table 1. Overview of main study procedure.

	Item
Pre-test	To assess general knowledge of each participant at baseline, the questions asking knowledge of the 20 scientific facts were presented as a modified CLOZE procedure. Questions were in a random order. Possible CLOZE responses were also presented in a random order. The following procedure was repeated for all 20 scientific facts. The order of fact presentation was randomised.
Fact presentation	One of the 20 scientific facts was presented in a random selection of one of the five variants (text alone, text accompanying cartoon, cartoon with explanatory agent, text accompanying photograph, and photograph with explanatory agent). Participants selected 'I have seen this image/read this text' to advance to the next screen.
Demographic measure and covariates	One of the segments of demographic or covariate measures requiring a response from participants. Possible solutions to English language proficiency CLOZE responses, and image options for Raven's progressive matrices, were presented in a random order.
Immediate recall	Questions about the scientific facts were presented in a modified CLOZE procedure, with possible responses presented in a random order.
Confidence Preference	A visual analogue scale directly below the immediate recall CLOZE response. After participants had completed the above procedure for all 20 facts, they selected which style they preferred from a list. The order of options was randomised.
Short-term recall	Questions about the 20 scientific facts were presented as a modified CLOZE procedure. Possible CLOZE responses were also presented in a random order.
Engagement	Participants self-rated how seriously they took the survey.

Reported time to complete is mean and standard deviation in brackets.

Data inclusion

Pre-test accuracy indicated participants were generally unfamiliar with the 20 facts (within participants $M = 5$, $SD = 3$, range 0 to 18 correct; for further detail see Supplemental Table S2). Removal of data for facts correctly answered at pre-test from the dataset resulted in a total of 4547 answers across the 310 participants for subsequent analysis, balanced well across presentation variants (text only $n = 908$, text caption with cartoon $n = 899$, text caption with photograph $n = 916$, cartoon with explanatory agent $n = 908$, and photograph with explanatory agent $n = 916$). There was no significant association between style and whether or not a given fact was excluded from analysis $\chi^2(4) = 2.51, p = 0.54$.

Hierarchical linear model results

A summary of hierarchical linear models can be seen in Figure 2, with full model coefficients reported in Supplemental Table S5. PEWST, cognitive-academic English language proficiency and Raven's progressive matrix scores were significantly positively associated with correct recall and confidence across all conditions.

There was no significant difference in immediate recall, short-term recall or confidence in answers related to whether the fact was presented as an image (rather than text), a cartoon (rather than a photo) or a text caption (rather than an explanatory agent). Immediate

recall was substantially higher than pre-test accuracy ($M = 18$, $SD = 2$, range 3–20 correct), falling slightly for short-term recall ($M = 17$, $SD = 3$, range 4–19 correct), demonstrating sufficient variability in learning and forgetting for subsequent analysis.

Participant preference for a particular presentation variant was, from most to least preferred: cartoon with explanatory agent (26%), 'I did not really prefer any option' (21%), 'Anything with a picture' (15%), text only (12%), text caption with cartoon (10%), text caption with a photograph (8%) and photograph with explanatory agent (8%). Model fit tests indicated assumptions were met for immediate and short-term recall (Supplemental Table S4 and Figure S1), and that models had conditional $R^2 \approx 0.6$ (able to detect medium to large effects, see Bosco, Aguinis).⁵⁹ Heavy skewing in confidence ratings effect ($M = 92$, $SD = 17$, range 0–100) led to a ceiling effect which destabilised aICC (0.01), model dispersion, and power (conditional $R^2 \approx 0.05$), indicating conclusions relating to confidence should be interpreted with caution.

Where preference for a particular presentation variant was not met (e.g. those facts presented in photos to a participant with a stated preference for text), immediate recall was statistically significantly worse (β ranging -3.89 to -0.54 , corresponding to up to four more facts recalled incorrectly), and confidence was significantly lower (β ranging -1.5 to -1.8 , corresponding to an approximately 2% drop in confidence)

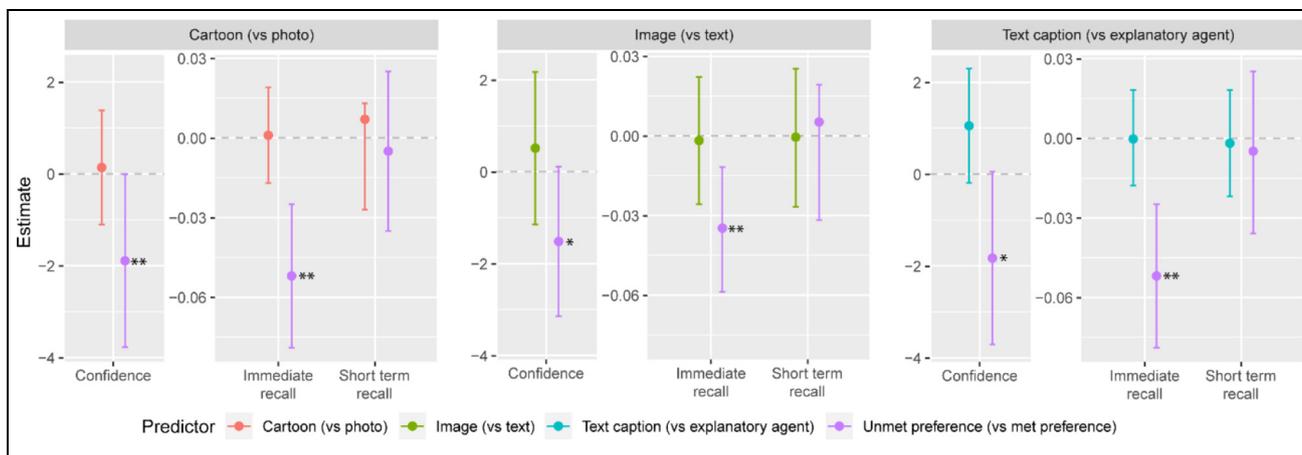


Figure 2. Summary of hierarchical linear model results.

Note. Values are fixed effect standardised coefficients for ease of comparison across all models. Error bars indicate 98% confidence intervals. * indicates significance at $\alpha < 0.05$, ** significance at $\alpha < 0.02$. All models controlled for PEW scientific literacy score, cognitive-academic English language proficiency score, and Raven's progressive matrix score. For tables including all coefficients and model fit indicates, see Supplemental Table S5.

than where preferences were met. Unmet preference was not significantly associated with short-term recall, and there were no significant differences in recall or confidence between those for whom preferences were met, and those who expressed no preference. All results were generally consistent across sensitivity analysis (Supplemental Tables S6–S8).

Discussion

This study investigated whether visual accompaniment of scientific information influences immediate and short-term fact recall in an adult audience and if that is influenced by different visual styles. We found no evidence of a particular benefit to scientific fact recall arising from any tested aspect of the 'manga guide to...' format; there was no systematic difference in fact recall depending on whether information was presented alongside a supporting image, whether that image was photograph or cartoon, and whether that image included an explanatory agent. This is distinct from several prior studies, which found a particular visual accompaniment improved information recall in comparison with text alone.^{35,60} We also did not find a clear preference for any particular presentation variant. However, alignment of visual presentation with participant's stated preference for presentation variant did significantly improve immediate and immediate fact recall. Our findings are supported by prior studies indicating that individuals both prefer and are more confident in technical information conveyed by visuals accompanied by text,⁶¹ and that confidence in information recall is positively associated with attitude to the style of information delivery.⁶²

Given prior evidence that immediate recall is more sensitive to contextual cues,³³ it was expected the different styles of visual accompaniment would influence immediate recall to a greater degree than short-term recall. The high accuracy and confidence at immediate recall, and subsequent lower accuracy at short-term recall suggests participants used working memory at immediate test. Following Baddeley's model of working memory,²⁶ engaging in such temporary mental rehearsal relies on cognitive processes that are more sensitive to the nature of distraction tasks than the procedure of information encoding, storage and retrieval inherent in recall. For example, a verbal distractor is less effective at disrupting a visual working memory task, but more effective at disrupting a verbal working memory task.²⁹ The phonological loop is implicated in sentence comprehension,⁶³ so is the most likely component of working memory participants used in the current study.

It is likely that the distraction tasks did not prevent participants using working memory because they were either insufficiently difficult (demographic, English language proficiency, scientific knowledge) or of a sufficiently different modality (the Raven's progressive matrices used here rely on the visuo-spatial sketchpad.⁶⁴ If this is the case, then results could be taken to suggest that the style of information presentation impacted working memory (rather than immediate recall), possibly due to the irrelevance of visuals to the capacity of the phonological loop. This account would align with previous findings that in the context of comics, people spend longer looking at text than illustrations.⁴⁷ Further, these results indicate that future work seeking to understand the impact of visual

accompaniment on scientific information my find it fruitful to either explicitly focus on working memory, or short-term recall rather than immediate recall.

Given the predominance of scholarly studies suggesting that visuals of varying kinds in combination with text should promote superior recall to text alone,^{6,16,35,65} it was surprising that we found no such effect. This discrepancy may be due to the predominance of research examining hypotheses generated by the dual coding theory, which comment on information presented across the modalities of symbolic/verbal and sensorimotor/visual. This paradigm implicitly focusses research attention on comparing two or three conditions only – text alone, visuals of some style and perhaps a pairing of text and visuals.^{6,16,35} In this study, we contrasted recall performance and preference across multiple combinations of text and different visual styles. This greater number of conditions allowed different spatial relationships between text and visuals across the styles (e.g. as a caption, or within a speech bubble). Generally, attention and recall is most pronounced when the text is presented as close as possible to the corresponding visual.⁶⁶ Eye-tracking studies from the advertising literature take this further, suggesting that text within a visual (e.g. a speech bubble) enhances attention to both visuals and text, possibly by physically pairing them in the visual field.⁶⁷ There is considerable scope for future examination of whether the physical proximity between text and visuals may alter how they are perceptually processed in relation to the dual coding theory.

In this study, the most preferred style was ‘cartoon with an explanatory agent’, this mirrors the real-world popularity of ‘manga guide to...’ format. Yet, results were unhelpful in disambiguating whether visual simplicity (cartoon versus photograph) or presence of an explanatory agent were more impactful for promoting recall because neither factor was unambiguously associated with recall performance. It is, however, useful to consider that the most popular visual style was not associated with better recall. This comports with a recent review of eleven studies, where Farinella (2018) noted no difference in scientific knowledge acquisition from text or comic book formats, but an almost universal preference for the comic format.⁴⁵ This supports the notion that irrelevant imagery, that is content that is not semantically related to the content, such as the inclusion of an explanatory agent, is preferable for reasons that do not directly benefit recall. It is unclear what these reasons may be, as non-substantive visual content such as narrative⁴⁸ and humour⁴⁹ have been linked with enhanced recall.

Prior research indicates marked individual differences in preferences and styles of learning,⁶⁸ and there is evidence this manifests in clear preferences for

particular styles of information presentation.¹ Yet, preferences in the current study were diffuse. The most popular – cartoon with an explanatory agent – was endorsed by just over a quarter of the participants, and a large number of participants indicated no preference for any visual style. If this reflects a genuine phenomenon, the reasons for this are unclear, beyond the extremely unlikely possibility that our sample by chance included a group with atypically heterogeneous preferences. It is instead more likely due to methodological differences between prior work and the current study. Unlike our findings, several studies have clearly indicated majority preference within their sample, ascertained by an ipsative paradigm where a preference must be stated.¹ Forcing a choice when there is genuinely no preference can lead to nonsensical responses that fit well into statistical models, but do not map onto the participant’s reality.⁶⁹ A useful model that future research could adopt is the preference rating in Moore et al.,¹⁶ which deconstructs preference into perception of varied aspects of a visualisation style (e.g. a series of questions rating factors such as ugly/beautiful on a scale of 1–5), or as in Luo,³⁶ which rates preference for each visual style separately on a scale of 1–7.

Aligning with both cognitive fit theory and the learning styles hypothesis, a variety of information presentation methods may be useful for recall of scientific facts in adults, the key factor being a match between information presentation and a particular individual’s predisposition for learning.^{40,41} While stated preference is not synonymous with learning predisposition, style, or preference, there is evidence that an individual’s preference for a particular style of visualisation is linked with cognitive style.³⁶ Causality remains unclear – it is possible people prefer a particular information presentation style because it facilitates memory, or conversely that it facilitates memory because they prefer it.

This study had a number of notable strengths, as well as limitations. To our knowledge this is the first examination of visual accompaniment of scientific information on recall in adults where participants were presented with multiple methods of combining text and images. This approach allows robust insight into within- and between- participant variability, resulting in findings that reflect individual characteristics that would be lost in simpler between-person analyses. Repeated pilot testing and exclusion of correctly answered facts at pre-test minimised bias from prior domain-specific knowledge of facts. A further strength was the design of the visual stimuli for comparison between photographs and illustration; the illustrations were developed directly from photographs, removing bias from mismatches in salient characteristics such as content, colour, and complexity. However, due to the focus on exploring elements of the ‘manga guide to...’

format, we did not explore whether specific visuals semantically reinforced or contextualised the presented facts. This lack of semantic connection may explain the lack of significant improvement in any recall when images were paired with text. Further, we did not ascertain whether the visual content of the styles was consistently perceived (e.g. abstracted cartoon clouds may not be as clearly cloud-like as a photograph, hence may prove a less salient cue for a fact relating to clouds). We also did not explore intermediate forms of abstraction in representation between photograph or illustration, or inherent information richness of visuals (as discussed in McCloud S and Manning¹⁵). Both of these shortcomings could be addressed in future work through the generation and pilot testing of stimuli according to some measure of semantic distance, considered here as the closeness between a visual and the meaning it holds. This could be achieved through the qualitative techniques outlined in Silvennoinen et al.⁷⁰

Unfortunately, variability within and between participant recall was somewhat limited by high accuracy at immediate and short-term test, which may have led to our missing trends within the data (type II error). Our focus on immediate and short-term memory has limited ecological validity for application in adult education settings, where long-term recall is of more interest. Similarly, due to how recall was tested, our results are only generalisable to rote memorisation via CLOZE procedures. Further, though comprehension relates to similar processes as recall (e.g. prior knowledge, mapping and memory constraints per McNamara et al.,¹⁹ the current study provides no insight into whether the presented facts were understood, rather than rote memorised.

Conclusions and next steps

In conclusion, we found that the style of visual accompaniment of scientific information in accordance with the ‘manga guide to...’ format influenced immediate, but not short-term, fact recall in an adult audience when written English literacy, scientific literacy and non-verbal intelligence were taken into account. Recall was enhanced when the participant’s preference for presentation variant was met, but there was no clear majority preference across participants. This indicates that short-term recall of scientific facts in adults may be best served by enquiring about, and then meeting participant preference for visual accompaniment. Future research should investigate how to proactively map out respondent preference, in order to most effectively use visuals and text for conveying scientific information.

Acknowledgements

The authors would like to thank Dr. Elizabeth Huxley for granting us permission to use her likeness as our ‘explanatory agent’, and patience in the photography session.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Erin Walsh  <https://orcid.org/0000-0001-8941-0046>

Supplemental material

Supplemental material for this article is available online.

References

1. Wang CH, Han D and Chou W-K. Overview of information visualisation in science education. *Int J Comput Sci Eng* 2018; 17(1): 118–125.
2. Bucchi M and Saracino B. “Visual science literacy” images and public understanding of science in the digital age. *Sci Commun* 2016; 38(6): 812–819.
3. Powell TE, Boomgaarden HG, De Swert K, et al. A clearer picture: the contribution of visuals and text to framing effects. *J Commun* 2015; 65(6): 997–1017.
4. Geise S and Baden C. Putting the image back into the frame: Modeling the linkage between visual communication and frame-processing theory. *Commun Theory* 2015; 25(1): 46–69.
5. Cook MP. Visual representations in science education: The influence of prior knowledge and cognitive load theory on instructional design principles. *Sci Educ* 2006; 90(6): 1073–1091.
6. Lin S-F and Lin H-S. Learning nanotechnology with texts and comics: the impacts on students of different achievement levels. *Int J Sci Educ* 2016; 38(8): 1373–1391.
7. Torres HN and Zeidler DL. The effects of English language proficiency and scientific reasoning skills on the acquisition of science content knowledge by Hispanic English language learners and native English language speaking students. *Electr J Res Sci Math Educ* 2002; 6(3): 1–59.
8. Wiebe EN, Clark AC and Hasse EV. Scientific visualisation: linking science and technology education through graphic communications. *J Des Technol Educ* 2001; 6(1): 40–47.
9. Stone C and Springer M. Interactivity, connectedness and ‘teacher-presence’: Engaging and retaining students online. *Aust J Adult Learn* 2019; 59(2): 146.
10. Rigutto C. The landscape of online visual communication of science. *J Sci Commun* 2017; 16(2): C06.

11. Leivas Pozzer L and Roth WM. Prevalence, function, and structure of photographs in high school biology textbooks. *J Res Sci Teach* 2003; 40(10): 1089–1114.
12. Vavra KL, Janjic-Watrich V, Loerke K, et al. Visualization in science education. *Alberta Sci Educ J* 2011; 41(1): 22–30.
13. Berger J. *Ways of seeing: based on the BBC television series*. British Broadcasting Corporation. London: Penguin Books, 1985.
14. Gilbert JK. Visualization: a metacognitive skill in science and science education. In: Gilbert JK (ed) *Visualization in science education*. Heidelberg, Germany: Springer, 2005, pp.9–27.
15. McCloud S and Manning A. Understanding comics: the invisible art. *IEEE Trans Prof Commun* 1998; 41(1): 66–69.
16. Moere AV, Tomitsch M, Wimmer C, et al. Evaluating the effect of style in information visualization. *IEEE Trans Vis Comput Graph* 2012; 18(12): 2739–2748.
17. Holt NJ, Bremner A, Sutherland E, et al. *Psychology: the science of mind and behaviour*. New York: McGraw-Hill Education, 2012.
18. Saranya N and Nandhini CU. Semi-supervised nonlinear distance metric learning via random forest and relative similarity algorithm. *Int Res J Eng Technol* 2016; 3(12): 912–917.
19. McNamara DS and Magliano J. Toward a comprehensive model of comprehension. *Psychol Learn Motiv* 2009; 51: 297–384.
20. Bobrow SA and Bower GH. Comprehension and recall of sentences. *J Exp Psychol* 1969; 80(3p1): 455.
21. Cowan N. What are the differences between long-term, short-term, and working memory? *Prog Brain Res* 2008; 169: 323–338.
22. Alloway TP and Gathercole S. Working memory and short-term sentence recall in young children. *Eur J Cogn Psychol* 2005; 17(2): 207–220.
23. Burgess N and Hitch G. Computational models of working memory: putting long-term memory into context. *Trends Cogn Sci* 2005; 9(11): 535–541.
24. Ranganath C, Johnson MK and D'Esposito M. Prefrontal activity associated with working memory and episodic long-term memory. *Neuropsychologia* 2003; 41(3): 378–389.
25. Barch DM, Csernansky JG, Conturo T, et al. Working and long-term memory deficits in schizophrenia: is there a common prefrontal mechanism? *J Abnorm Psychol* 2002; 111(3): 478.
26. Baddeley AD and Hitch G. *Working memory. The psychology of learning and motivation*. New York, NY: Academic, 1974.
27. Diamond A. Executive functions. *Annu Rev Psychol* 2013; 64: 135–168.
28. Hitch GJ. Temporal grouping effects in immediate recall: a working memory analysis. *Q J Exp Psychol Sec A* 1996; 49(1): 116–139.
29. Logie RH, Zucco GM and Baddeley AD. Interference with visual short-term memory. *Acta Psychol* 1990; 75(1): 55–74.
30. Bauer PJ. Long-term recall memory: behavioral and neuro-developmental changes in the first 2 years of life. *Curr Dir Psychol Sci* 2002; 11(4): 137–141.
31. Ward PJ and Walker JJ. The influence of study methods and knowledge processing on academic success and long-term recall of anatomy learning by first-year veterinary students. *Anat Sci Educ* 2008; 1(2): 68–74.
32. Potter MC and Lombardi L. Regeneration in the short-term recall of sentences. *J Mem Lang* 1990; 29(6): 633–654.
33. Hopper WJ and Huber DE. Learning to recall: Examining recall latencies to test an intra-item learning theory of testing effects. *J Mem Lang* 2018; 102: 1–15.
34. Wickelgren WA. Associative intrusions in short-term recall. *J Exp Psychol* 1966; 72(6): 853.
35. Anglin GJ and Stevens JT. Prose-relevant pictures and recall from science text. *Percept Mot Ski* 1986; 63(3): 1143–1148.
36. Luo W. User choice of interactive data visualization format: The effects of cognitive style and spatial ability. *Decis Support Syst* 2019; 122: 113061.
37. Cummins J. Cognitive/academic language proficiency, linguistic interdependence, the optimum age question and some other matters. Working Papers on Bilingualism, No. 19. 1979.
38. Das JP, Mishra RK and Kirby JR. Cognitive patterns of children with dyslexia: A comparison between groups with high and average nonverbal intelligence. *J Learn Disabil* 1994; 27(4): 235–242.
39. Raven J. The Raven's progressive matrices: change and stability over culture and time. *Cogn Psychol* 2000; 41(1): 1–48.
40. Vessey I, Zhang P and Galletta D. The theory of cognitive fit. Human-computer interaction and management information systems. *Foundations* 2006; 5: 141–183.
41. Cuevas J. An analysis of current evidence supporting two alternate learning models: learning styles and dual coding. *J Educ Sci Psychol* 2016; 6(1): 1–13.
42. Norman D. Emotion & design: attractive things work better. *Interactions* 2002; 9(4): 36–42.
43. Evans MA, Watson C and Willows DM. A naturalistic inquiry into illustrations in instructional textbooks. In: Houghton HA and Willows DM (eds) *The psychology of illustration*. New York: Springer, 1987. pp.86–115.
44. Gurung RA and Martin RC. Predicting textbook reading: the textbook assessment and usage scale. *Teach Psychol* 2011; 38(1): 22–28.
45. Farinella M. The potential of comics in science communication. *J Sci Commun* 2018; 17(01): Y01.
46. Rentoul AJ and Fraser BJ. Predicting learning from classroom individualization and actual-preferred congruence. *Stud Educ Evaluation* 1980; 6(3): 265–277.
47. Bucher HJ and Boy B. How informative are information comics in science communication? Empirical results from an eye-tracking study and knowledge testing. In: Dunst ALJ, Laubrock J and Wildfeuer J (eds) *Empirical comics research*. New York: Routledge, 2018, pp.176–196.
48. Dettori G and Paiva A. Narrative learning in technology-enhanced environments. In: Duval E, Sharples M and

- Sutherland R (eds) *Technology-enhanced learning*. Cham, Switzerland: Springer, 2009, pp.55–69.
49. Skalski P, Tamborini R, Glazer E, et al. Effects of humor on presence and recall of persuasive messages. *Commun Q* 2009; 57(2): 136–153.
 50. Park S and Lim J. Promoting positive emotion in multimedia learning using visual illustrations. *J Educ Multimedia Hypermedia* 2007; 16(2): 141–162.
 51. Kennepohl D and Roesky HW. Drawing attention with chemistry cartoons. *J Chem Educ* 2008; 85(10): 1355.
 52. Lin YY, Holmqvist K, Miyoshi K, et al. Effects of detailed illustrations on science learning: an eye-tracking study. *Instr Sci* 2017; 45(5): 557–581.
 53. Porter D. Modified cloze procedure: a more valid reading comprehension test. *Engl Lang Teach J* 1976; 30(2): 151–155.
 54. Kostyk A, Zhou W and Hyman MR. Using surveytainment to counter declining survey data quality. *J Bus Res* 2019; 95: 211–219.
 55. Caulfield J and Smith WC. The reduced redundancy test and the cloze procedure as measures of global language proficiency. *Mod Lang J* 1981; 65(1): 54–58.
 56. PEW. Public's knowledge of science and technology. 2003. <https://www.pewresearch.org/politics/2013/04/22/publics-knowledge-of-science-and-technology/>
 57. Bates D, Sarkar D, Bates MD, et al. The lme4 package. *R Package Version* 2007; 2(1): 74.
 58. Wu H and Leung S-O. Can Likert scales be treated as interval scales?—a simulation study. *J Soc Serv Res* 2017; 43(4): 527–532.
 59. Bosco FA, Aguinis H, Singh K, et al. Correlational effect size benchmarks. *J Appl Psychol* 2015; 100(2): 431.
 60. Wang Z, Wang S, Farinella M, et al. Comparing effectiveness and engagement of data comics and infographics. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, Glasgow, United Kingdom; 2019.
 61. Kools M, van de Wiel MW, Ruiter RA, et al. Pictures and text in instructions for medical devices: effects on recall and actual performance. *Patient Educ Couns* 2006; 64(1-3): 104–111.
 62. Lee YH. Manipulating ad message involvement through information expectancy: effects on attitude evaluation and confidence. *J Advert* 2000; 29(2): 29–43.
 63. Lauro LJR, Reis J, Cohen LG, et al. A case for the involvement of phonological loop in sentence comprehension. *Neuropsychologia* 2010; 48(14): 4003–4011.
 64. DeShon RP, Chan D and Weissbein DA. Verbal overshadowing effects on Raven's advanced progressive matrices: evidence for multidimensional performance determinants. *Intelligence* 1995; 21(2): 135–155.
 65. Houts PS, Doak CC, Doak LG, et al. The role of pictures in improving health communication: a review of research on attention, comprehension, recall, and adherence. *Patient Educ Couns* 2006; 61(2): 173–190.
 66. Mayer RE. The promise of multimedia learning: using the same instructional design methods across different media. *Learn Instruct* 2003; 13(2): 125–139.
 67. Boerman SC, Smit EG and van Meurs L. Attention battle; the abilities of brand, visual, and text characteristics of the ad to draw attention versus the diverting power of the direct magazine context. In: Okazaki S (ed) *Advances in advertising research* (vol 2). Wiesbaden, Germany: Springer, 2011, pp.295–310.
 68. Loo* R. Kolb's learning styles and learning preferences: is there a linkage? *Educ Psychol* 2004; 24(1): 99–108.
 69. O'Mahony M and Wichchukit S. The evolution of paired preference tests from forced choice to the use of 'No Preference' options, from preference frequencies to d' values, from placebo pairs to signal detection. *Trends Food Sci Technol* 2017; 66: 146–152.
 70. Silvennoinen JM, Kujala T and Jokinen JP. Semantic distance as a critical factor in icon design for in-car infotainment systems. *Appl Ergon* 2017; 65: 369–381.