Learning and Instruction 80 (2022) 101633

Contents lists available at ScienceDirect

Learning and Instruction

journal homepage: www.elsevier.com/locate/learninstruc

Mental imagery skill predicts adults' reading performance

Sebastian Suggate [a,*], Wolfgang Lenhard [b]

[a] Department of Educational Science, University of Regensburg, Universitätsstraße 31, 93053, Regensburg, Germany
[b] Department of Psychology IV, University of Wuerzburg, Wittelsbacherplatz 1, 97074, Würzburg, Germany

ARTICLE INFO

Keywords:
Mental imagery
Mental simulation
Situational models
Reading speed
Reading comprehension

ABSTRACT

Mental imagery is foundational to human experience, lying at the heart of cognition and reading, however research has failed to conclusively investigate and demonstrate a link. Therefore, we conducted three studies measuring adults' reading and imagery performance. In Study 1, the mental imagery skills of 155 adults were measured using two established self-report measures, namely the Plymouth Sensory Imagery Questionnaire (Psi-Q) and the Spontaneous Use of Imagery Scale (SUIS), and a novel imagery comparison task. In Study 2 ($n = 452$), a control for speeded processing replaced the SUIS. In Study 3 ($n = 236$), we added a measure of reading speed. Findings indicate that the objective measurement of mental imagery was associated with reading performance, whereas self-report measures were not. Further, reading comprehension linked more strongly to mental imagery than reading speed did. Findings demonstrate, for the first time, that mental imagery processes are intrinsically linked with reading performance.

One of the most pleasant goals of leisure reading is to be completely immersed in a captivating story and be transported into a fictitious reality, where we can sympathize with the protagonist, experience dangerous adventures, and gain experiences that are elusive to us in real life (Johnson et al., 2013). Alongside emotional engagement, transportation, and attention, mental imagery appears to be key for immersion (Mak et al., 2020) and simulation of seemingly real experiences from text (Leopold et al., 2019). However, in previous reading research, mental imagery has only played an, at best, minor role, being subservient to decoding and comprehension factors (National Reading Panel, 2000). Therefore, across three studies, we test whether mental imagery and reading skill are linked in adult readers, while examining propositional versus imagery-related reading, reading speed, and reading comprehension.

1. Mental imagery

Mental imagery refers to the experience of sensory stimuli in the absence of their physical presence or perception (Kosslyn, 1994). Thus, mental images represent private events of a quasi-perceptual nature (Kosslyn et al., 2010). Because mental images mostly represent perceptual experiences, they can pertain to any of the sensory modalities, namely the visual, tactile, auditory, olfaction, gustation, proprioception, and the vestibular senses (Andrade et al., 2014; Borst, 2013), with vision, followed by hearing, being the most studied (Finke & Slayton, 1988; Malouin et al., 2009). Further, research using fMRI methods has demonstrated that during mental imagery, a similar neural activation to that observed during perception occurs (Ehrsson et al., 2003; Kosslyn et al., 2001). Accordingly, mental imagery can be conceived of as a shadow of perception (Kosslyn et al., 2010). In light of theories of grounded cognition, according to which perception and experience are closely intertwined (Barsalou, 1999), mental imagery takes on a new importance as a fundamental underlying skill for cognition and learning (Martzog & Suggate, 2019; Suggate & Martzog, 2021).

Given that mental imagery often involves representing the perceptual world internally, it can be conceived of as a form of mental simulation, involving constructing a mental model of external events and stimuli (Zwaan, 1999). Fundamental imagery processes include generating, inspecting, manipulating, and transforming images (Kosslyn et al., 1990). Mental imagery has been measured using a variety of tasks that often involve participants viewing then visually scanning internal models of previously viewed stimuli, mental rotation tasks, or self-report questionnaires (Pylyshyn, 2002). Due to the reliance on memory inherent in many of these tasks, it is not surprising that some have suggested great overlap between imagery and memory processes (Addis, 2020). Indeed, as we later argue, the lack of a robust mental imagery task involving an objective and a psychometrically stable measurement

* Corresponding author.
E-mail addresses: sebastian.suggate@ur.de (S. Suggate), wolfgang.lenhard@uni-wuerzburg.de (W. Lenhard).

https://doi.org/10.1016/j.learninstruc.2022.101633
Received 15 September 2021; Received in revised form 18 May 2022; Accepted 19 May 2022
Available online 26 May 2022

across many items has hampered investigating the role of mental imagery in psychological processes.

1.1. Reading skill

Skilled readers can read (most) texts effortlessly and fluently, applying technical prowess at decoding symbols into a content that is meaningful for the reader (Cain & Oakhill, 2011; Keenan et al., 2008). Regarding the technical aspects, skilled reading relies on reading fluency and reading comprehension, with reading fluency requiring rapid naming (Kirby et al., 2003), alphabet knowledge, knowledge of print conventions, and decoding skill (Joshi et al., 2012; National Reading Panel, 2000; Snow et al., 1998). Besides decoding, fundamental processes involved in reading comprehension include phonological, syntactic, and semantic aspects of language (e.g., Suggate et al., 2018), inference making, strategy usage, vocabulary and prior knowledge (Cromley & Azevedo, 2007). Importantly, however, engaging in the practice of reading occurs with the purpose not only of developing these technical aspects (e.g., increasing vocabulary), but to gain information or enrich mental experience of places and states without having to physically be present. To this end, reading can involve the creation of mental models for the content depicted in the text, thereby forming a crucial aspect of the reading experience (Zwaan, 1999; Zwaan & Madden, 2004).

1.2. The role of mental imagery in comprehension

Reading fundamentally involves converting code into a conscious experience extending beyond the symbols that comprise that code. Consequently, comprehension has been proposed to comprise the construction of *mental models* (Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991) or *situation models* (Van Dijk & Kintsch, 1983), which refer to the complex representations that can arise during reading to depict the scenes and content conveyed by the text. Interestingly, the philosopher Wittgenstein (1922) already stated that reasoning is not based on logical processes, but on the ability to generate mental images. Experimental research underlines that during reading perceptual representations and features, such as shape or color, are routinely activated (e.g., Richter & Zwaan, 2010). Further, brain regions activated during word and sentence comprehension overlap or border brain regions activated by actions (Hargreaves et al., 2012; Hauk et al., 2001; Pulvermüller et al., 2001). Accordingly, it is conceivable that mental images represent the perceptual units used and experienced in comprehension processes.

However, it is generally agreed that the experience in consciousness arising from reading manifests in (at least) two modalities, namely a verbal and a perceptual code (Paivio, 2013; Sadoski & Paivio, 2013). The verbal code can be thought of as propositional, comprising abstract amodal statements. The perceptual code involves the creation of mental images that are experienced as having perceptual qualities across a range of sensory modalities (Gentsch et al., 2016). Importantly, it appears that readers switch between codes during reading, with some passages evoking vivid imagery, whereas other passages or genres are processed in a verbal code (Kuzmičová, 2014).

Kuzmičová (2014) identified, partly based on research and partly based on introspection, four different kinds of mental imagery occurring during reading. Briefly, and according to her terminology, enactment imagery occurs spontaneously and creates the experience of actually being in the story. Description imagery involves more of a view from outside, often involving visual representations of scenes. Two further types of imagery represent the experience of a voice reading the text (i.e., speech) or one enacting a verbal exchange (i.e., rehearsal). In short, it appears that reading, at least some of the time and for most people, involves the creation of conscious mental imagery, which leads to the question of how mental imagery affects reading performance.

1.3. Empirical findings relating to imagery and reading

Empirical studies on the connections between imagery and reading can be categorized into three groups. The first set uses designs involving some form of a contradiction effect, in which participants read text that implies an activity in one modality (e.g., someone speaking). In a subsequent step, a sentence or word might appear that is either consistent or discordant with imagery evoked by the first sentence, participants are required to judge the consistency, and response latencies are recorded. Findings across a range of methods and measures indicate generally that participants construct mental models containing modality-specific mental imagery (Clement & Falmagne, 1986; Fincher-Kiefer, 2001; Gunraj et al., 2014; Klin & Drumm, 2010; Kurby et al., 2009; Pilotti et al., 2000).

A second set of studies has looked at how using different imagery strategies affects reading comprehension, with visualization long belonging to the repertoire of reading comprehension interventions (e.g., Author, 2010). Asking children to use mental imagery strategies improves reading performance (De Koning et al., 2017; e.g.,; Pressley, 1976 but cf.; Sadoski et al., 1990; Mccallum & Moore, 1999) also for non-fiction STEM texts ($d > .60$, Leopold et al., 2019). Further, visualization strategies have been found to improve memorization ($\eta^2 = 0.325$, Marre et al., 2021) and empathy during reading of fiction texts ($d = 0.92$, Johnson et al., 2013). However, it could also be argued that some non-specified feature of imagery strategy use such as attention resulted in the gains, or that the imagery strategies resulted in increased propositional processing. Accordingly, these studies alone do not definitively demonstrate that imagery skill improves reading.

Third, there are a small number of studies that have tested whether mental imagery skill relates to reading skill. This is an important line of investigation, providing a stringent test of the idea that imagery capacities relate to reading comprehension. Specifically, by measuring mental imagery and reading skill using performance tests separated in time, as opposed to studies relying on the contradiction effect, an independent estimate of the influence of the one domain on the other can be calculated because each is measured offline from the other. At a practical level, such research also informs whether mental imagery is a valuable skill in its own right and that consequently, it could even be a valuable goal to target in educational settings.

1.4. Current studies

Introspective reports (Kuzmičová, 2014) and theories of reading (Zwaan, 1999) posit that readers experience, at least some of the time, mental images in the mental models they create while reading. These images appear to be comprised of information from multiple sensory modalities (Pulvermüller et al., 2001). Furthermore, neuropsychological investigations indicate that during language processing activation of sensorimotor networks can occur. If reading occurs in either a multimodal imagery and/or a verbal code, this gives rise to two possibilities. In the first, mental imagery is important for reading of texts that require or support the development of mental models but not for texts requiring only abstract, propositional coding. In a second instance, given that perceptual systems may underlie language processing—even if this is not consciously experienced as imagery—mental imagery may relate to both imagery and propositional reading experiences.

However, previous research is scarce and methodologically problematic regarding the question of whether mental imagery skill links to reading performance, and also as to whether such links are specific to processing in a non-verbal code. Some studies rely on self-report imagery measures ($r = 0.26$ to 0.34, Boerma et al., 2016; Mol et al., 2016), which, given that imagery is a private event, would appear to have questionable validity (e.g., how am I to reliably rate whether I am better or worse at imagining a specific object than anyone else?). Three studies from one research group have investigated links between mental imagery and reading performance (Commodari et al., 2020; Guarnera et al., 2017, 2019). However, the imagery tasks employed showed inconclusive links and were conceptually and methodologically problematic (e.g., ceiling effects). For instance, one task was better conceptualized as a measure of memory because children reproduced symbols from cards after 15sec of exposure and another appeared to measure cross-modal transfer (i.e., haptic exploration of objects). The remaining task involved imagery for letters, which would expectedly be better solved by children with greater reading skill due to their more advanced letter knowledge. Effect sizes for the letter imagery tasks were generally significant predictors of reading performance in one study ($\beta = 0.24$, Guarnera et al., 2017), but not in the others (Commodari et al., 2020; Guarnera et al., 2019).

Accordingly, research is needed to test for links between mental imagery skill and reading performance in an offline paradigm (i.e., both tasks measured separately from one another). Second, such research needs to include reading measures that require mental model generation and those that are more heavily rooted in a verbal, propositional code. Third, research needs to employ a performance oriented mental imagery measure that is not conceptually confounded with reading skill. A recent development of a mental imagery task appears to satisfy these criteria (Martzog & Suggate, 2019). Specifically, this task assesses mental imagery using a mental comparison task in which participants draw forth two perceptual images and are asked to compare these. For instance, in a haptic-visual item, participants respond to a question "what is sharper [stimulus property], a nail [stimulus 1] or a needle [stimulus 2]?" The stimuli are chosen to be objectively verifiable, such that it is unlikely that responses can be provided based on propositional knowledge alone. Further, it is unlikely that participants have a series of propositions embedded in declarative knowledge for such obscure comparisons as whether a violin is shinier than a trumpet. Response accuracy and latency are recorded. The task has proven successful in measuring mental imagery in children (Martzog & Suggate, 2019; Suggate & Martzog, 2021) and here we apply this task to an adult sample across three studies to investigate, for the first time using objective skill-focused measures, whether mental imagery skill relates to reading performance.

2. Study 1

Although mental models posit that mental imagery plays an important role in reading comprehension (Zwaan, 1999), research is lacking on whether mental imagery skill is associated with reading performance. Further, previous work has either relied on self-report imagery tasks (Boerma et al., 2916; Mol et al., 2016), or used mental imagery measures that are confounded by letter-knowledge, memory, or cross-modal transfer (Commodari et al., 2020; Guarnera et al., 2017, 2019). Additionally, previous work has not compared links between mental imagery and reading for texts that differ in terms of whether they are classified as being imagery heavy or verbal.

As a first step to determining whether mental imagery relates to reading performance, we recruited 155 adults, thus extending previous research to a population of accomplished readers. A key reason for this decision is that we could remove any developmental effects expected in child samples, where both reading (Juel, 1988) and imagery skills (Isaac & Marks, 1994) undergo rapid change. Additionally, if offline mental imagery plays a role in reading performance, as would be predicted, then the first step to establishing this would be to test this on samples of proficient readers. For this reason, we recruited university students studying education. To test our supposition that self-report imagery measures may be unsatisfactory predictors, we included an objective measure, namely the imagery comparison task (ICT; Martzog & Suggate, 2019), and two self-report measures, namely the Spontaneous Use of Imagery Scale (SUIS; Görgen et al., 2016) and the Plymouth Sensory Imagery Questionnaire (Psi-Q; Andrade et al., 2014). To measure reading performance, we incorporated a sentence verification task (SVT). One key advantage of the SVT was that this allows for an easy manipulation of mental imagery, by including sentences that either require the construction of mental images or that are heavily propositional. We hypothesized that:

(a) ICT performance would positively predict SVT performance,
(b) ICT should be a stronger predictor of sentences that were rated as requiring greater mental imagery, as evidenced by a ICT × imagery interaction, and
(c) given the difficulty inherent in self-report imagery scales, we did not expect the Psi-Q or the SUIS to predict SVT performance.

2.1. Method

Participants. Participants were 155 University students who were invited to participate as part of an introductory education lecture for mostly first year students. As outlined later, data for six participants scoring below chance were dropped, leaving 149. The students were aged between 18 and 55 years ($M = 22.84$, $SD = 5.56$) and 99% were born in Germany. Due to an experimenter error, gender was not recorded so was derived from subjects' first names, whereby an estimated 80% were female. Additionally, 17% had already received a higher education degree. We estimated the effects size for linear mixed models that could be detected with this sample size via the sjstats package (Lüdecke, 2021; function '*samplesize_mixed*'). Given the number of observations and cases, a medium effect size of $f^2 = 0.2$ could be detected with a moderate power of $1 - \beta = 0.80$.

Measures. At the beginning of the study, subjects provided demographic data (i.e., country of birth, school and higher education qualifications, age).

Reading skill. Reading performance was measured using the SVT, with this type of task having been validated as a measure of reading performance in adult readers (Lüdtke et al., 2019). In the SVT, participants read a series of 123 sentences varying in length from 24 to 80 characters ($M = 47.01$, $SD = 12.07$) and have to decide, as promptly as possible, whether the information contained in the sentence was "true" or "false". We designed the task such that 61 items referred to factual sentences (e.g., "the euro is a currency"), designed to tap more strongly into propositional knowledge, and others that required the generation of mental images (e.g., "frogs are smaller than ants"). Each item was rated by seven graduate research assistants, none of whom were involved in data collection, according to the extent that during reading the sentences they experienced mental images or abstract concepts. Responses were recorded on a seven-point Likert scale (1 = clear mental images, 7 = only abstract concepts). The ratings showed a good distribution with both imagery and abstract items present, $M = 3.86$, $SD = 2.05$.

The computer recorded response latencies and accuracy. Six items were dropped because participants responded close to chance level, defined as an accuracy of less than 70%. Internal consistency derived from data from Studies 1 and 2 indicated excellent internal consistency for response accuracy, $\alpha_{cr} = 0.98$, and excellent for response latency, $\alpha_{cr} = 0.97$.

Plymouth Sensory Imagery Questionnaire (Psi-Q). The Psi-Q is a self-report measure of the clarity of mental images in seven sensory domains, namely vision, hearing, tactile, taste, olfaction, body, and emotional imagery (Andrade et al., 2014). Five questions pertain to each modality, each of which is answered on an 11-point Likert scale, ranging from "no mental image" to "clear and alive, like in real life". For the current study, given our lack of hypotheses regarding separate imagery modalities, we created a combined imagery score out of the sum of all items. In support of this decision, the internal consistency was exceptionally high, $\alpha_{cr} = 0.90$.

Spontaneous Use of Imagery Scale (SUIS). The SUIS was used as a second self-report measure of mental imagery (Görgen et al., 2016). The SUIS is a 12-item scale measuring participants' experience of imagery as this naturally appears in everyday experience. In the German adaptation six further items were added (Görgen et al., 2016). Participants respond on a 5-point Likert scale from "never" to "always" to statements such as "When I listen to the news, real-life images appear in my mind". The scale has been shown to be unidimensional (Görgen et al., 2016) and correlates with, for example, the Psi-Q (Andrade et al., 2014). In the current study, the internal consistency was high, $\alpha_{cr} = 0.70$.

Imagery comparisons task (ICT). An ICT was used to measure participants' mental imagery (Moyer, 1973; Paivio, 1975). This task has recently been adapted and further developed (Martzog & Suggate, 2019). In this task, participants are presented with a series of words and are asked to make judgments (visual and/or haptic) based on a sensory property of the stimuli's corresponding mental images. Previous research indicates that the ICT is sensitive to individual differences in children's fine motor skills (Martzog & Suggate, 2019) and screen-media usage (Author, 2020), thus given that this task depends on response latencies, there is good reason to suspect that it will function well with adults. The stimuli were presented auditorily to avoid any confound with reading skill.

Specifically, participants were asked to imagine two specific objects, and then requested to make a judgment as to which from the target and distractor item was better encapsulated by a sensory feature (i.e., "which is shinier, [a] trumpet or [a] violin?"). Thus, the question "which is" was played at the beginning of the trial (0 s), followed by the adjective (1 s), then the first target imagery item (2 s), "or" (4 s), and then the second target imagery item (5 s). The presentation of each target imagery item was accompanied with a marker (i.e., a small square and the corresponding key press) on the left and right sides of the screen, to serve as a reminder as to which key press was paired with which stimulus. Participants were instructed to respond as quickly and accurately as possible by pressing the "f" key for the first target item and the "j" key for the second. Note that the question was thus phrased, such that the target imagery stimuli appear at the end of the sentence, so that the comparison can only begin after presentation of the final stimuli. In addition, in German, the indefinite article "a" was not grammatically necessary in the question sentence, thus reducing memory load between presentation of the two target stimuli.

Response accuracy and latency were both recorded and in total there were 80 ICT items. Response latency was conceptualized as being the key dependent variable, particularly in adult samples, given that it was expected that participants would be able to answer close to all items correctly. Further, response latency should better reflect the mental imagery systems proficiency at readily creating mental imagery. Item order was randomized for each individual participant. Internal consistency for response accuracy was $\alpha_{cr} = 0.94$ and McDonald's $\omega_{cr} = 0.96$ (computed with the R package *psych*; Revelle, 2018) for both winsorized and nonwinorized response latency.

Procedure. The study was conducted as an online correlational study, with mental imagery conceptualized as a predictor variable and reading performance as an outcome variable.

The study was programmed in PsychoPy 2020.1.3 and run in Pavlovia (Peirce et al., 2019), which has demonstrated timing accuracy across a range of browsers of less than 3.5 msec (Bridges et al., 2020). Participants were invited to take part via clicking on the link and received detailed instructions as to what they would require, including headphones, a laptop or PC with the screen positioned 40–50 cm from the participant's head. Participants first answered demographic questions, and then the Psi-Q, SUIS and the mental imagery task, followed by questions on reading habits (not reported here), the ICT, and finally the reading performance was measured with the SVT. In total, the study lasted between 20 and 30 min. Participation was voluntary but encouraged to provide students experience of experimental research as part of course participation. University Ethics procedures were followed, which in this case according to University Ethics procedures did not require a full ethical review due to the anonymous and voluntary participation and with the kind of data collected being non-sensitive.

Data analyses. Data were first screened for outliers and skew and kurtosis were checked, descriptives and correlation coefficients were calculated. To test the influence of mental imagery on reading performance, mixed effect linear models (MELM) were used. MELM represent a state-of-the-art method for modelling response latency performance, accounting for item level variance in response accuracy while testing the effects of predictors (Raudenbush & Bryk, 2002). The advantage of this procedure is that in a task with dual demands, namely, to respond as quickly AND as accurately as possible, response latency and accuracy cannot be treated independently. One solution is to exclude all inaccurate responses; however, this approach is problematic because in a task with two response alternatives, even 50% of unknown responses will be correctly guessed.

Accordingly, we modelled individual response latency and accuracy at an item level for each of the SVT responses (level 1). Including individual intercepts at level 2 accounts for subject differences in response latency (Raudenbush & Bryk, 2002). Further level 1 predictors included item features, such as the length of the sentences in the SVT and whether the item was imagery versus propositional (dummy coded, 0 = propositional, 1 = imagery). Analyses were conducted in R version 4.1.0 using the lmerTools (Kuznetsova et al., 2017) and lme4 packages (Bates et al., 2015). Aside from the MELM and homogeneity indicators, analyses were conducted in SPSS 26. The code for all studies, data, and analyses are open-source and upon publication will be made available (Suggate, 2022).

2.2. Results and discussion for study 1

Participants' reading performance scores were winzorized at three standard deviations above or below their individual means, with a maximum lower score of 0.5 s. Only 0.02% of responses in total were below this threshold and 1.44% of data at the upper end were above this and hence capped at the individual upper bound. For the ICT, one extreme outlier was removed (2800 s) then individual means, SD, and upper bounds at 3 standard deviations above the individual mean were calculated. Using this threshold, 1.73% of data were capped at the upper individual mean. Further, six participants performed at lower than chance on the ICT, so these were removed from all analyses.

Descriptive statistics were calculated for the reading and imagery measures, and these appear in Table 1. Inspecting kurtosis and skewness statistics revealed that data were generally normally distributed (all values below 2), but that the response latencies for reading performance and ICT, as well as accuracy for ICT were slightly peaked. Given MELMs robustness against violations of normality, the data were not transformed. As can be seen in Table 1, the ICT correlated the strongest with reading out of the imagery measures ($r = 0.31, p < .01$). Accuracy on the ICT positively correlated with both the SUIS and the PsiQ (both at $r = 0.18$, $p < .05$).

To test the contributions of the imagery measures for reading performance, MELM were conducted with two levels as random intercepts, namely the participant level and a nested item level describing features of the reading test items. Thus, response latency on the reading performance task was modelled using maximum likelihood with random effects at the item level and at the participant level. Predictors at the item level included whether the sentences referred to an imagery item and sentence length (number of characters). Predictors at the participant level included the three imagery measures. The model is presented in Table 2.

As can be seen in Table 2, ICT performance significantly predicted reading performance, $p < .001$. Sentence length was also a significant predictor, $p < .001$, but neither the SUIS nor Psi-Q was, nor their interactions with the imagery rating.

In Study 1, we found that mental imagery was a predictor of reading performance, as hypothesized. Additionally, we found that self-report questionnaires did not significantly relate to reading performance, consistent with the idea that an objective measure of mental imagery, such as the ICT, is needed to become a reliable contributor to explaining reading performance. To our knowledge, this was the first study to

Table 1
Descriptive statistics and correlation coefficients for the imagery and reading measures from study 1.

	Measure	Min.	Max.	*M*	*SD*	1	2	3	4	5	6
1	SVT (sec)	.651	7.34	2.45	1.14	–	-.07	.31**	-.15	-.11	-.17*
2	SVT (% acc)	78.00	96.00	89.34	3.00		–	.06	.33**	.14	.14
3	ICT (sec)	.48	2.94	1.26	.46			–	-.00	.19*	-.04
4	ICT (% acc)	63.00	98.00	88.38	5.42				–	.18*	.18*
5	PsiQ	4.49	9.00	6.95	.97					–	.16*
6	SUIS	1.83	4.78	3.45	.53						–

Note. n = 149, SVT = sentence verification task, ICT = imagery comparisons task, SUIS = Spontaneous Use of Imagery Scale, Psi-Q = Plymouth Sensory Imagery Questionnaire.

Table 2
Mixed effects linear model predicting reading performance from mental imagery in study 1.

Predictors	Reading performance (SVT, sec)		
	Estimates	CI	p
Intercept	.85	-.66–2.36	.269
Item level			
Reading (% acc)	-.67	-.77–-.56	<.001
SVT imagery rating	-.13	-.31–.05	.157
Sentence length (characters)	.05	.04–.06	<.001
Subject level			
ICT (sec)	.99	.64–1.34	<.001
PsiQ	-.08	-.25–.09	.352
SUIS	-.02	-.32–.28	.898
Cross level interaction			
SVT imagery rating X ICT	.02	-.02–.06	.428
SVT imagery rating X PsiQ	.01	-.01–.03	.403
SVT imagery rating X SUIS	.01	-.02–.05	.539
Random Effects			
σ^2	2.78		
$\tau_{00\ date}$	.74		
$\tau_{00SVT\_rating\_loop.thisIndex}$	.23		
ICC	.26		
N_{date}	149		
$N_{SVT\_rating\_loop.thisIndex}$	117		
Observations	17433		
Marginal R^2/Conditional R^2	.357		

demonstrate, using a clear and objective measure of mental imagery, that the latter is meaningfully related to reading performance.

3. Study 2

Given the novelty of the finding in Study 1, that mental imagery links to reading performance in adult readers, it bears replication and extension, which we now turn to. Regarding extension, it could be argued that both the SVT and ICT measures both had a similar format, involving rapid and accurate responses, whereas the SUIS and Psi-Q were untimed. Accordingly, a concern arises that the ICT and SVT were related more strongly because they both capitalize on processing speed or some underlying cognitive or task demand. Consequently, including a third, similarly constructed measure that is not targeted at mental imagery should control for commonalities between the SVT and ICT measures. Thereby, we can be more certain that any remaining links between the ICT and SVT are not due to some unidentified confound.

Thus, we replicated Study 1 with a larger sample and included a measure to control for task format. Specifically, we included an auditory processing task that follows the exact same format as the ICT, but which does not require mental image comparisons, but instead requires participants to decide which of two sounds is the louder. As such, this task is closely matched to the ICT, with the exception that mental images are not compared, and hence operates as a stringent control measure. Additionally, we dropped the SUIS from Study 2 for two reasons. First, the Psi-Q seemed the slightly stronger predictor in Table 2, and second, the Psi-Q measures the clarity of mental imagery, not its spontaneous appearance, which seemed more relevant for reading performance.

We hypothesized that the ICT would predict reading performance, above and beyond the remaining predictors and including the sound comparisons task.

3.1. Method

Participants. Participants were 452 university students aged between 17 and 38 years (M = 20.06, SD = 2.52). Ninety-nine percent were born in Germany, 79% were female, and 21% already had a university degree.

Procedure and measures. The measures and design were similar to Study 1 with the exception that the SUIS was not administered, an auditory processing task was added, and the ICT was shortened. In terms of procedure, the current data were collected in an online study that was pre-registered and where all code and data will be published (Suggate & Martzog, 2021). At the beginning of the study, a pretest containing 16 ICT and 8 auditory processing items was administered, which are reported here. Homogeneity again was high with ω_{cr} = 0.91 for the response latency data of the ICT. Thus, compared to Study 1, mental imagery was measured using a shortened form.

The auditory processing task was added to provide a processing speed measure, to discount the possibility that the links found between mental imagery and reading performance in Study 1 were due to both of these tasks having a speed component. The auditory processing task was conceptualized to run precisely the same as the ICT, with the exception that the comparison pertained to the perceived loudness of two beeps. In terms of underlying mental processes, the auditory processing task requires memory for the loudness of sounds, as opposed to imagery generation and inspection. Thus, participants were asked "which is louder [beep1] or [beep2]?" The loudness of the beeps was varied, such that there were five different volume variations of the same beep of approximately 500msec in length. If participants took longer than 5 s to respond, the next sound was presented.

3.2. Results and discussion for study 2

The same procedure to winsorize the data as in Study 1 was used. For the reading performance task, this resulted in 0.60% of the data being capped at .5 s, and 1.44% of the data being capped at three standard deviations above the participants' individual means. For the ICT, only 0.12% of data were raised to the lower bound of 0.50 s, and only 0.15% of data at three standard deviations above the individual means. For the auditory processing task, there were no outliers. The descriptive statistics and correlation coefficients between the reading performance, mental imagery, and auditory processing tasks were calculated and appear in Table 3. The auditory processing task correlated with the ICT, $r = 0.25, p < .01$, as did the latter with reading performance, $r = 0.25, p < .01$. Inspection of kurtosis and skewness statistics revealed similar distributions to that in Study 1.

A MELM, similar to that in Study 1, was conducted to test the unique contribution that mental imagery made to reading performance, above and beyond the auditory processing task. The model is presented in Table 4. As can be seen in Table 4, the auditory processing task explained significant variance in reading performance, $p < .01$, but

Table 3
Descriptive statistics and correlation coefficients for the imagery and reading measures from study 2.

	Measure	Min.	Max.	*M*	*SD*	1	2	3	4	5	6	7
1	SVT (sec)	.50	16.20	2.31	1.45	–	.07	.25**	-.01	.00	.09	-.04
2	SVT (% acc)	.07	.97	.86	.16		–	-.05	.07	.07	.02	-.05
3	ICT (sec)	.46	8.03	1.82	.84			–	-.16**	.01	.25**	-.06
4	ICT (% acc)	.31	1.00	.82	.11				–	.07	-.06	.14**
5	PsiQ	2.74	9.00	6.74	1.06					–	.05	-.06
6	Auditory processing (sec)	.51	3.90	1.63	.56						–	-.25**
7	Auditory processing (% acc)	.13	1.00	.89	.14							–

Note. $n = 452$, SVT = sentence verification task, ICT = imagery comparisons task, Psi-Q = Plymouth Sensory Imagery Questionnaire.

Table 4
Mixed effects linear model predicting reading performance from mental imagery in study 2.

Predictors	Reading performance (SVT, sec)		
	Estimates	CI	p
Intercept	−0.27	−1.00–0.46	0.467
Item level			
SVT (% acc)	.65	.29–1.02	<.001
SVT imagery rating	−0.01	−0.10–0.08	0.811
Sentence length (characters)	.05	.04–.06	<.001
Subject level			
ICT (sec)	.28	.19–.36	<.001
PsiQ	-.03	-.09–.03	.347
Auditory processing (sec)	.15	.04–.26	.007
Cross level interaction			
SVT imagery rating X ICT	.00	-.01–.01	.442
SVT imagery rating X PsiQ	.00	-.01–.01	.590
Random Effects			
σ^2	1.60		
$\tau_{00\ subject}$	.40		
$\tau_{00\ item}$	.30		
ICC	.31		
$N_{subject}$	452		
N_{item}	117		
Observations	52884		
Conditional R^2	.412		

Note. SVT = sentence verification task, ICT = imagery comparisons task, Psi-Q = Plymouth Sensory Imagery Questionnaire.

mental imagery as measured by the ICT remained significant, $p < .001$. The self-report Psi-Q imagery measure again did not uniquely predict reading performance, $p = .35$. Also, the ICT did not interact with SVT imagery rating, indicating that mental imagery did not play a greater role in predicting reading of imagery versus conceptual items.

The findings from Study 2 replicate and extend those of Study 1 by confirming the finding that mental imagery, when recorded with an objective measure, links to reading performance in adult readers. Further, the findings discount the idea that a common processing or task requirement inherent in both the ICT and the reading measure led to spurious findings, by virtue of our including an auditory processing control variable. Despite the large sample size, we again found that the Psi-Q did not link to reading performance, supporting our supposition that mental imagery research requires measures that are objective. One limitation of both Study 1 and 2 is that we included only one measure of reading, based on a SVT. This measure contains both fluency and comprehension components; however, it may prove useful to try to measure fluency separately. It would be expected that fluency links less strongly to mental imagery than a comprehension measure, because mental models are not strictly required when reading with a focus on fluency.

4. Study 3

Studies 1 and 2 showed that mental imagery performance, as measured by the ICT, links to reading performance. The reading performance in both prior studies contained both a fluency and a comprehension component, because participants achieved better scores for their prompt reading of the sentences and clicking on the correct response. In Study 3, we include a more pure reading speed measure, with corresponding reading comprehension questions. Thereby, we tested the contribution of mental imagery to reading speed in comparison to reading comprehension and the SVT from Studies 1 and 2. We reasoned that reading speed would be less strongly related to the ICT because this does not necessitate the formation of mental imagery. Accordingly, we hypothesized that:

a) ICT would predict reading comprehension more strongly than reading speed, and
b) that ICT would predict SVT even when controlling for reading speed.

4.1. Method

Participants. The participants were 236 students aged 20.74 years ($SD = 2.55$) from an undergraduate lecture in education, 63% were female, 97% were born in Germany, and 18% already had a higher education degree.

Measures and procedure. As with the previous studies, data were collected via an online correlational study programmed in Psychopy by the first author. The data from Study 3 were partly drawn from an online pre-registered study testing the effect of viewing film clips versus reading texts on mental imagery performance (Suggate & Martzog, 2021). Briefly, in the original experiment, participants viewed various film clips or read texts for 1 min each and after each viewing/reading they performed a block six items of the ICT. These film clips and texts had small effects (approx. 10 msec) on the initial response latencies for the first 1–3 items in each block. Given that the focus here is on inter-individual predictors of reading, the small effect of the conditions should not compromise the validity of the current experiment, especially because all participants performed the same conditions. There were also eight ICT trials at the beginning of the study, giving 80 items in total, thus matching the number of trials in Study 1. Reading performance was measured using the SVT from Studies 1 and 2 and was administered at the end of the study.

Reading speed was estimated based on how many sentences of connected text participants read within 1 min. Specifically, texts from modern adult thriller fiction were selected (Dan Brown's *Inferno*, Trudi Canavan's *The Rogue*, and Ian McEwan's *Nutshell*). We selected this genre to increase the chances of creating immersion and hence mental imagery. The texts comprised coherent passages of between 596 and 682 words ($M = 629$, $SD = 39$) and had a Flesh-Kincaid Grade Reading score of between grades 7.1 and 8.5. These were then divided into individual sentences, on average 24 per story ($SD = 5.76$). Participants were asked to read the text at their normal reading speed and were informed that they would be answering questions on these texts afterwards. This step was added to ensure that participants read the texts carefully, but quickly. Texts were presented one sentence at a time to each participant, who then pressed the space bar to indicate when they had finished reading the sentence and the next one was presented. The number of sentences that participants read in 1 min was recorded.

To control for participants actually paying attention to the stories, a number of reading comprehension questions were included after the entire block of stories and corresponding ICTs. The comprehension questions referred to the passages and required simple TRUE/FALSE responses. Given that we expected the students to read, on average, about half of the sentences, we ensured that two comprehension questions were drawn from the first third of the texts, two from the middle, and one question from the latter part. Questions referred to details mentioned in the story (e.g., "Cery was the leader of the gang") and can thus be conceived of as relatively challenging. Internal consistency for this scale was low, $\alpha_{cr} = .20$, so items were dropped if they had an at or near chance response accuracy defined as below 60% response accuracy. In total, 14 items were dropped which improved the internal consistency, $\alpha_{cr} = .46$. Thus, two scores were provided: (a) reading speed, defined as the total number of sentences read and with a theoretical range of between zero and 144, and (b) reading comprehension, representing the number of questions correctly answered, between zero and 16.

4.2. Results and discussion study 3

Following a similar procedure to Studies 1 and 2, only 0.17% of the reading performance data were winsorized. One further item with a lower than 70% response accuracy in the reading task was dropped. For the ICT data, three participants had a number of response latencies that were high (e.g., above 10sec) which was inflating their overall response latencies, hence these were also capped at 10sec (i.e., 58 in total). Similarly, implausibly fast responses (30 in total) were capped at .15sec before winsorizing. The descriptive statistics and correlation coefficients were calculated and appear in Table 5. As Table 5 shows, the reading measures correlated with one another, $|r|$s between 0.13 and 0.29, all ps $< .05$. Mental imagery was again correlated with reading performance, $p < .001$. Since reading comprehension and fluency was uncorrelated, $r = -0.12$, $p > .05$, there was no indication of a speed accuracy trade off.

To test the hypothesis that ICT would be a stronger predictor of reading comprehension than reading speed, simple regression analyses were conducted, given that data was not available in long (nested) format. Analyses indicated that faster responses on the ICT predicted better reading comprehension scores, $\beta = -0.23$, $p < .001$, $R^2 = 06$, and reading fluency, $\beta = -0.17$, $p = .01$, $R^2 = 03$, with the difference in betas not being statistically significant. To test the link between ICT and SVT controlling for reading speed, a MELM identical to those in Studies 1 and 2 was conducted, with the exception that reading speed was added as a predictor. This analysis is presented in Table 6. As with the earlier studies, ICT was a significant predictor of SVT and reading speed contributed significantly.

The findings from Study 3 replicate and extend those from the previous two studies. Specifically, we again found that mental imagery, as measured by the ICT, was a significant predictor of reading performance, even after controlling for reading speed. This suggests that our SVT likely tapped reading comprehension skill, explaining variance over and above fluency. Additionally, this indicates that mental imagery is likely a foundation skill for reading comprehension, as posited by theories of embodied cognition and mental models (Zwaan & Madden, 2004). The imagery rating was again not significant, indicating that mental imagery predicted abstract conceptual reading as much as it did high imagery reading. However, given that our reading comprehension measure assessed memory for explicit, perhaps propositional, details in the story, it might not have been optimally suited to tap participants' mental models. Accordingly, we may have underestimated links between mental imagery and reading comprehension.

Table 6
Mixed effects linear model predicting reading performance from mental imagery in study 3.

Predictors	Reading (SVT, sec)		
	Estimates	CI	p
Intercept	.58	-.02–1.17	.057
Item level			
SVT (% acc)	-.40	-.46– −.33	<.001
SVT imagery rating	.00	-.07–.08	.906
Sentence length (characters)	.05	.04–.05	<.001
Subject level			
ICT (sec)	.83	.63–1.03	<.001
Reading speed	-.01	-.01–-.01	<.001
Cross level interaction			
SVT imagery rating X ICT	-.01	-.03–.02	.636
Random Effects			
σ^2	1.50		
$\tau_{00\ date}$	.42		
$\tau_{00\ Sent\_Ver\_Loop.thisIndex}$	.26		
ICC	.31		
N_{date}	236		
$N_{Sent\_Ver\_Loop.thisIndex}$	116		
Observations	27376		
Marginal R^2/Conditional R^2	.445		

Note. SVT = sentence verification task, ICT = imagery comparisons task.

Further, we found some evidence that mental imagery was a significant predictor of reading speed also. This suggests that the mental imagery system might be involved in text processing, even if participants are not conscious of forming mental images or mental models. Although reading speed is often measured by counting the number of words per minute, our measure represented the number of sentences read over the course of six separate 1-min intervals. The reason for measuring reading speed at the sentence level arose purely due to the difficulty in measuring at a word level in an online study. However, our measure showed criterion validity with the other reading measures and good variance in the absence of floor and ceiling effects.

4.3. General discussion

The current study demonstrates, in our view and to our knowledge, for the first time that when using an objective measure, mental imagery uniquely related to reading comprehension. Over the course of three studies, with a total of 837 participants, we found links with reading performance after controlling for processing speed, measure type (self-report vs. objective), and reading skill (i.e., comprehension vs. fluency). Accordingly, our findings support the idea that mental imagery forms one key foundation of the process of mental model construction (Zwaan, 1999; Zwaan & Madden, 2004). Thereby, we were able to address concerns arising out of previous studies that may not have optimally operationalized mental imagery (Commodari et al., 2020; Guarnera

Table 5
Descriptive statistics and correlation coefficients for the imagery and reading measures from study 3.

	Measure		Min.	Max.	*M*	*SD*	1	2	3	4	5	6
1	SVT (sec)	236	1.07	7.71	2.76	1.35	–	-.12	.36**	-.04	-.15*	-.13*
2	SVT (% acc)	236	.61	.96	.88	.04		–	-.04	.47**	-.28**	.30**
3	ICT (sec)	236	.33	4.75	1.50	.47			–	-.11	-.10	-.22**
4	ICT (% acc)	236	.45	.96	.87	.07				–	-.20**	.18**
5	Reading speed (sentences)	236	8	144	47.58	21.00					–	-.29**
6	Reading comprehension	236	5	16	11.37	2.36						–

Note. SVT = sentence verification task, ICT = imagery comparisons task.

et al., 2017, 2019) or relied on self-report (Boerma et al., 2016).

Further, in all studies we manipulated the text read in the SVT so that the sentences invoked mental images or abstract conceptual knowledge. This idea is based on theories of reading and language processing in which readers may process the text as an abstract verbal code or they may find themselves immersed in a world of images (Kuzmičová, 2014; Paivio, 1975; 2013). As far as conscious imagery is concerned, it would have been expected that mental imagery should be particularly important for reading of high imagery sentences. However, this was not the case as evidenced by the lack of interactions between ICT and SVT item imagery ratings. On the other hand, some theories posit that even if not conscious, perceptual systems underlie both mental imagery and reading/language processing (Author, 2019; Hauk et al., 2001; Pulvermüller et al., 2001; Zwaan, 1999). The current findings support the latter assumption.

Unfortunately, but somewhat expectantly and in line with our suppositions, our findings did little to support the use of self-report scales, such as the PsiQ and the SUIS, in measuring mental processes to a sufficient degree of precision required for experimental paradigms. Specifically, despite the large samples employed here, the PsiQ did not predict reading performance. Post-hoc exploratory analyses in which the PsiQ scales were divided into the seven imagery modalities that they purportedly measure were unsatisfactory. In the first instance, such a division results in only five items per modality, which may be insufficient to adequately capture and sample imagery. Second, these subscales generally did not predict reading performance when entered into models. Accordingly, although self-report measures certainly have value in given settings (e.g., helping patients in clinical settings identify relative strengths in imagery, Andrade et al., 2014), they did not explain a significant amount of variance here.

5. Limitations

It is important to note that the current design is correlational, hence findings may be confounded by third variables. We tried to identify and rule out plausible confounds and the results provide a strong indication that mental imagery facilitates reading comprehension. This is, however, no definitive causal proof. Plausibly, it could be that reading comprehension skill improves mental imagery skill. Therefore, the current set of investigations could be furthered by including other controls typical in reading research, such as working memory, rapid naming and mental speed (Kirby et al., 2003). Although including these variables would be important, we accounted for processing speed in Study 2 and also to an extent in Study 3 by including reading speed. The finding from Study 3 that mental imagery accuracy, not response latency, predicted reading performance when controlling for reading speed indicates that mental imagery is linked to reading beyond the influence of processing speed.

5.1. *Directions for future research*

To identify the causal relationship between imagery and comprehension, future work could employ experimental designs, whereby the visuo-spatial sketchpad could be suppressed through an interference task. Additionally, future work could use the ICT to test links between mental imagery and cognitive skills, such as aspects of intelligence and language development. The ICT has also been used successfully with preschool and school-age children (Author, 2009), hence future research could study mental imagery in relation to reading development in younger samples.

We believe that the ICT employed here has considerable potential to advance the field of mental imagery research. One key advantage of the task is that it provides an objective measure, because, across a population and with a sufficient number of items, it is generally possible to arrive at stimuli that are objectively comparable and hence verifiable. For example, although there are certainly violins that are shinier than some trumpets, even a well-maintained violin glistens and blinds to an inferior extent to that of a freshly polished crimson-tinged golden member of the brass family. Additionally, using the ICT it is theoretically possible to test a range of modalities, extending beyond the haptic and visual domains tested here. Obvious candidates are the auditory (e.g., louder, shriller, deeper), kinesthetic (e.g., further), and gustatory (e.g., saltier, sweeter) modalities. Again, the items would have to be constructed in a way that comparisons cannot be solved through retrieval of facts from long-term memory, but instead elicit mental imagery processes. Such work could test for individual differences, including relative strengths, for specific imagery abilities in both clinical and educational settings (e.g., autism spectrum disorder, pica, schizophrenia).

Indeed a key feature of the current studies was that we found links between mental imagery and reading performance measured in an offline manner. Thereby we extend previous work that manipulated mental models using online paradigms (Clement & Falmagne, 1986; Fincher-Kiefer, 2001; Gunraj et al., 2014; Klin & Drumm, 2010; Kurby et al., 2009; Pilotti et al., 2000). The theoretical significance of this point is twofold. First, if two processes are linked in an offline paradigm, it may provide insight into commonalities underlying both (Anderson, 2007; Glenberg et al., 2008), as has been found with motor skills and mental imagery using an offline paradigm (e.g., Martzog & Suggate, 2019).

Second, if mental imagery skills are meaningfully linked to reading comprehension, then it could be a promising approach for educators to foster imagery abilities in children, not only in the context of comprehension strategies that draw on visualization, but as well as a general cognitive skill. Interventions promoting mental imagery would of course require further experimental support and development, given the correlational nature of our results. However, there is support for this notion, both from a conceptual (De Koning & van der Schoot, 2013) as well as an empirical point of view (Mak et al., 2020). Since short term effects from the manipulation in experimental studies are less pronounced than individual traits with respect to imagery, fostering these traits would have to begin early in education and extend over a long period of time. For instance, before reading skill is established, sharing stories with children may constitute one avenue as these involve mental model creation (Nyhout & O'Neill, 2017).

Currently, educators often prioritize propositional knowledge in the form of discrete informational units, such as facts or specific academic skills (Suggate, 2015). These aspects are extremely important of course, but from a grounded cognition perspective, it might be argued that direct sensory experiences form the foundation of concepts (Barsalou, 1999) and hence should not be disregarded. Thus, it may be counter-productive to depart prematurely from direct experience of the world, as is typically evident in early childhood play (Levine et al., 2012; Wallace & Russ, 2015), to academic and propositional ways of learning, as is becoming increasingly common in preschools and schools (Marcon, 2012). In any case, it would appear that there may be nothing to lose in stimulating mental imagery, even if the effect on reading comprehension turns out to be small or if it is limited to reading comprehension tasks requiring imagination. Viewing mental imagery as a skill to be acquired alongside other skills, such as reading and arithmetic, could enrich the educational landscape, thereby returning the experience of reading to the skill of reading.

Finally, although the contribution of this paper is fundamentally theoretical because of the pioneering nature of this work, we permit ourselves to conclude with some speculation. Taking Dual Code Theory (Paivio, 1975; 2013), Kuzmičová's (2014) distinctions, and findings from cognitive science around egocentric and allocentric perceptual processing (Paillard, 1991), we suggest that imagery might be best represented via two dimensions. Concerning the first dimension, we suggest a first versus third person perspective, whereby one is either in the scene and "experiencing" phenomena directly, on one is "perceiving" the scene from outside. These experiences are typically perceptual in nature, containing voices, images, smells, feelings of movement etc. In the second dimension, the symbolism level may be represented, being either a sensory versus propositional/verbal code. Concerning the latter, there has been much debate as to the nature of mental symbols, with some arguing that all mental symbols are embodied or perceptual in nature (e.g., Lakoff & Johnson, 2010) and others arguing that they are abstract and symbolic (Pylyshyn, 2002). However, most accept that it can be both, and that it likely varies from person to person, as do we accept this. Adding a further layer to the puzzle, it is unclear whether consciousness is a necessary or sufficient condition for mental imagery: if the mental imagery system is stimulated during symbolic processing but participants do not consciously experience mental images, is mental imagery important for this processing?

Author statement

The first author programmed the experiments and both authors, conceptualized, prepared and revised the manuscript.

Data availability statement and declaration of interests

The data of the study, the code for the experiment and analyses will be freely available on OSF (when published). There are no conflicts of interest that may have influenced the planning, conduct, and analysis of the study.

References

Addis, D. R. (2020). Mental time travel? A neurocognitive model of event simulation. *Review of Philosophy and Psychology, 11*(2), 233–259. https://doi.org/10.1007/s13164-020-00470-0

Anderson, M. L. (2007). The massive redeployment hypothesis and the functional topography of the brain. *Philosophical Psychology, 20*(2), 143–174. https://doi.org/10.1080/09515080701197163

Andrade, J., May, J., Deeprose, C., Baugh, S.-J., & Ganis, G. (2014). Assessing vividness of mental imagery: The plymouth sensory imagery questionnaire. *British Journal of Psychology, 105*(4), 547–563. https://doi.org/10.1111/bjop.12050

Author, 2010; 2015; 2018; 2019; 2021.

Barsalou, L. W. (1999). Perceptions of perceptual symbols. *Behavioral and Brain Sciences, 22*(4), 637–660. https://doi.org/10.1017/S0140525X99532147

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Boerma, I. E., Mol, S. E., & Jolles, J. (2016). Reading pictures for story comprehension requires mental imagery skills. *Frontiers in Psychology, 7*, 1630. https://doi.org/10.3389/fpsyg.2016.01630

Borst, G. (2013). Neural underpinnings of object mental imagery, spatial imagery, and motor imagery. In K. N. Ochsner, & S. Kosslyn (Eds.), *The oxford handbook of cognitive neuroscience* (Vol. 2). Oxford University Press.

Bridges, D., Pitiot, A., MacAskill, M. R., & Peirce, J. W. (2020). The timing mega-study: Comparing a range of experiment generators, both lab-based and online. *PeerJ, 8*, Article e9414. https://doi.org/10.7717/peerj.9414

Cain, K., & Oakhill, J. (2011). Matthew effects in young readers: Reading comprehension and reading experience aid vocabulary development. *Journal of Learning Disabilities, 44*, 431–443.

Clement, C. A., & Falmagne, R. J. (1986). Logical reasoning, world knowledge, and mental imagery: Interconnections in cognitive processes. *Memory & Cognition, 14*(4), 299–307. https://doi.org/10.3758/bf03202507

Commodari, E., Guarnera, M., Di Stefano, A., & Di Nuovo, S. (2020). Children learn to read: How visual analysis and mental imagery contribute to the reading performances at different stages of reading acquisition. *Journal of Psycholinguistic Research, 49*(1), 59–72. https://doi.org/10.1007/s10936-019-09671-w

Cromley, J. G., & Azevedo, R. (2007). Testing and refining the direct and inferential mediation model of reading comprehension. *Journal of Educational Psychology, 99*(2), 311–325. https://doi.org/10.1037/0022-0663.99.2.311

de De Koning, B. B., Bos, L. T., Wassenburg, S. I., & van der Schoot, M. (2017). Effects of a reading strategy training aimed at improving mental simulation in primary school children. *Educational Psychology Review, 29*(4), 869–889. https://doi.org/10.1007/s10648-016-9380-4

De Koning, B. B., & van der Schoot, M. (2013). Becoming part of the story! Refueling the interest in visualization strategies for reading comprehension. *Educational Psychology Review, 25*(2), 261–287. https://doi.org/10.1007/s10648-013-9222-6

Ehrsson, H. H., Geyer, S., & Naito, E. (2003). Imagery of voluntary movement of fingers, toes, and tongue activates corresponding body-part-specific motor representations. *Journal of Neurophysiology, 90*(5), 3304–3316. https://doi.org/10.1152/jn.01113.2002

Fincher-Kiefer, R. (2001). Perceptual components of situation models. *Memory & Cognition, 29*(2), 336–343. https://doi.org/10.3758/BF03194928

Finke, R. A., & Slayton, K. (1988). Explorations of creative visual synthesis in mental imagery. *Memory & Cognition, 16*(3), 252–257. https://doi.org/10.3758/bf03197758

Gentsch, A., Weber, A., Synofzik, M., Vosgerau, G., & Schütz-Bosbach, S. (2016). Towards a common framework of grounded action cognition: Relating motor control, perception and cognition. *Cognition, 146*, 81–89.

Glenberg, A. M., Sato, M., Cattaneo, L., Riggio, L., Palumbo, D., & Buccino, G. (2008). Processing abstract language modulates motor system activity. *Quarterly Journal of Experimental Psychology, 61*(6), 905–919. https://doi.org/10.1080/17470210701625550

Görgen, S. M., Hiller, W., & Witthöft, M. (2016). Die spontaneous use of imagery scale (suis)–entwicklung und teststatistische prüfung einer deutschen adaption. *Diagnostica, 62*(1), 31–43. https://doi.org/10.1026/0012-1924/a000135

Guarnera, M., Faraci, P., Commodari, E., & Buccheri, S. L. (2017). Mental imagery and school readiness. *Psychological Reports, 120*(6), 1058–1077. https://doi.org/10.1177/0033294117717262

Guarnera, M., Pellerone, M., Commodari, E., Valenti, G. D., & Buccheri, S. L. (2019). Mental images and school learning: A longitudinal study on children. *Frontiers in Psychology, 10*, 2034. https://doi.org/10.3389/fpsyg.2019.02034

Gunraj, D. N., Drumm-Hewitt, A. M., & Klin, C. M. (2014). Embodiment during reading: Simulating a story character's linguistic actions. *Journal of Experimental Psychology Learning Memory and Cognition, 40*(2), 364–375. https://doi.org/10.1037/a0034853

Hargreaves, I. S., Leonard, G. A., Pexman, P. M., Pittman, D. J., Siakaluk, P. D., & Goodyear, B. G. (2012). The neural correlates of the body-object interaction effect in semantic processing. *Frontiers in Human Neuroscience, 6*. https://doi.org/10.3389/fnhum.2012.00022

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2001). Somatotopic representation of action words in human motor and premotor cortex. *Neuron, 41*, 301–307.

Isaac, A. R., & Marks, D. F. (1994). Individual differences in mental imagery experience: Developmental changes and specialization. *British Journal of Psychology, 85*(4), 479–500. https://doi.org/10.1111/j.2044-8295.1994.tb02536.x

Johnson-Laird, P. N. (1983). Mental models. In *Towards a cognitive science of language, inference and consciousness*. Cambridge, UK: Cambridge University Press.

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Essays in cognitive psychology*. Lawrence Erlbaum Associates, Inc.

Johnson, D. R., Cushman, G. K., Borden, L. A., & McCune, M. S. (2013). Potentiating empathic growth: Generating imagery while reading fiction increases empathy and prosocial behavior. *Psychology of Aesthetics, Creativity, and the Arts, 7*(3), 306–312. https://doi.org/10.1037/a0033261

Joshi, R. M., Tao, S., Aaron, P. G., & Quiroz, B. (2012). Cognitive component of componential model of reading applied to different orthographies. *Journal of Learning Disabilities, 45*(5), 480–486. https://doi.org/10.1177/0022219411432690

Juel, C. (1988). Learning to reading and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology, 80*, 437–447.

Keenan, J. M., Betjemann, R. S., & Olson, R. K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading, 12*, 281–300.

Kirby, J. R., Parrila, R. K., & Pfeiffer, S. L. (2003). Naming speed and phonological awareness as predictors of reading development. *Journal of Educational Psychology, 95*, 453–464.

Klin, C. M., & Drumm, A. M. (2010). Seeing what they read and hearing what they say: Readers' representation of the story characters' world. *Psychonomic Bulletin & Review, 17*(2), 231–236. https://doi.org/10.3758/PBR.17.2.231

Kosslyn, S. M. (1994). *The resolution of the imagery debate*. MIT Press.

Kosslyn, S. M., Ganis, G., & Thompson, W. L. (2001). Neural foundations of imagery. *Nature Reviews Neuroscience, 2*, 635–642.

Kosslyn, S. M., Margolis, J. A., Barrett, A. M., Goldknopf, E. J., & Daly, P. F. (1990). Age differences in imagery abilities. *Child Development, 61*(4), 995. https://doi.org/10.2307/1130871

Kosslyn, S. M., Thompson, W. L., & Ganis, G. (2010). The case for mental imagery. In *Oxford psychology series* (Vol. 39)Oxford University Press.

Kurby, C. A., Magliano, J. P., & Rapp, D. N. (2009). Those voices in your head: Activation of auditory images during reading. *Cognition, 112*(3), 457–461. https://doi.org/10.1016/j.cognition.2009.05.007

Kuzmičová, A. (2014). Literary narrative and mental imagery: A view from embodied cognition. *Style, 48*(3), 275–293. https://doi.org/10.13140/RG.2.2.30990.77128

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lakoff, G., & Johnson, M. (2010). *Philosophy in the flesh: The embodied mind and its challenge to Western thought*. Basic Books.

Leopold, C., Mayer, R. E., & Dutke, S. (2019). The power of imagination and perspective in learning from science text. *Journal of Educational Psychology, 111*(5), 793–808. https://doi.org/10.1037/edu0000310

Levine, S. C., Ratliff, K. R., Huttenlocher, J., & Cannon, J. (2012). Early puzzle play: A predictor of preschoolers' spatial transformation skill. *Developmental Psychology, 48*(2), 530–542. https://doi.org/10.1037/a0025913

Lüdecke, D. (2021). sjPlot: Data visualization for statistics in social science. The comprehensive R network. available: https://CRAN.R-project.org/package=sjPlot.

Lüdtke, J., Froehlich, E., Jacobs, A. M., & Hutzler, F. (2019). The sls-berlin: Validation of a German computer-based screening test to measure reading proficiency in early and late adulthood. *Frontiers in Psychology, 10*, 1682. https://doi.org/10.3389/fpsyg.2019.01682

Mak, M., Vries, C. D., & Willems, R. M. (2020). The influence of mental imagery instructions and personality characteristics on reading experiences. *Collabra: Psychology, 6*(1), 43. https://doi.org/10.1525/collabra.281

Malouin, F., Richards, C. L., Durand, A., Descent, M., Poiré, D., Frémont, P., Pelet, S., Gresset, J., & Doyon, J. (2009). Effects of practice, visual loss, limb amputation, and disuse on motor imagery vividness. *Neurorehabilitation and Neural Repair, 23*(5), 449–463. https://doi.org/10.1177/1545968308328733

Marcon, R. A. (2012). The importance of balance in early childhood programs. In S. P. Suggate, & E. Reese (Eds.), *Contemporary debates in child development and education* (pp. 159–168). Routledge, Taylor & Francis.

Marre, Q., Huet, N., & Labeye, E. (2021). Embodied mental imagery improves memory. *Quarterly Journal of Experimental Psychology, 74*(8), 1396–1405. https://doi.org/10.1177/17470218211009227

Martzog, P., & Suggate, S. P. (2019). Fine motor skills and mental imagery: Is it all in the mind? *Journal of Experimental Child Psychology, 186*, 59–72. https://doi.org/10.1016/j.jecp.2019.05.002.

Mccallum, R. D., & Moore, S. (1999). Not all imagery is created equal: The role of imagery in the comprehension of main ideas in exposition. *Reading Psychology, 20* (1), 21–60. https://doi.org/10.1080/027027199278493

Mol, S. E., Jolles, J., van Batenburg-Eddes, T., & Bult, M. K. (2016). Early adolescents' and their parents' mental imagery in relation to perceived reading competence. *Journal of Research in Reading, 39*(3), 253–267. https://doi.org/10.1111/1467-9817.12045

Moyer, R. S. (1973). Comparing objects in memory: Evidence suggesting an internal psychophysics. *Perception & Psychophysics, 13*(2), 180–184. https://doi.org/10.3758/BF03214124

National Reading Panel. (2000). In *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. National Institute of Child Health and Development. http://www.nichd.nih.gov/publications/nrp/report.htm.

Nyhout, A., & O'Neill, D. K. (2017). Children's enactment of characters' movements: A novel measure of spatial situation model representations and indicator of comprehension. *Mind, Brain, and Education, 11*(3), 112–120. https://doi.org/10.1111/mbe.12142

Paillard, J. (1991). Motor and representational framing of space. In J. Paillard (Ed.), *Brain and space* (pp. 163–182). Oxford University Press.

Paivio, A. (1975). Perceptual comparisons through the mind's eye. *Memory & Cognition, 3*(6), 635–647. https://doi.org/10.3758/BF03198229

Paivio, A. (2013). *Mind and its evolution: A dual coding theoretical approach*. Psychology Press Taylor & Francis Group.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). Psychopy2: Experiments in behavior made easy. *Behavior Research Methods, 51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Pilotti, M., Gallo, D. A., & Roediger, H. L. (2000). Effects of hearing words, imaging hearing words, and reading on auditory implicit and explicit memory tests. *Memory & Cognition, 28*(8), 1406–1418. https://doi.org/10.3758/bf03211841

Pressley, G. M. (1976). Mental imagery helps eight-year-olds remember what they read. *Journal of Educational Psychology, 68*(3), 355–359. https://doi.org/10.1037/0022-0663.68.3.355

Pulvermüller, F., Harle, M., & Hummel, F. (2001). Walking or talking? Behavioral and neurophysiological correlates of action verb processing. *Brain and Language, 78*(2), 143–168. https://doi.org/10.1006/brln.2000.2390

Pylyshyn, Z. W. (2002). Mental imagery: In search of a theory. *Behavioral and Brain Sciences, 25*, 157–238.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods (2nd)*. Sage Publications.

Revelle, W. (2018). *psych: Procedures for personality and psychological research*. Evanston, Illinois, USA: Northwestern University. https://CRAN.R-project.org/package=psych.

Richter, T., & Zwaan, R. A. (2010). Integration of perceptual information in word access. *Quarterly Journal of Experimental Psychology, 63*(1), 81–107. https://doi.org/10.1080/17470210902829563, 2006.

Sadoski, M., Goetz, E. T., Olivarez, A., Lee, S., & Roberts, N. M. (1990). Imagination in story reading: The role of imagery, verbal recall, story analysis, and processing levels. *Journal of Reading Behavior, 22*(1), 55–70. https://doi.org/10.1080/10862969009547694

Sadoski, M., & Paivio, A. (2013). *Imagery and text: A dual coding theory of reading and writing* (2). Routledge.

Snow, C. E., Burns, M. S., & Griffin, P. (1998). *Preventing reading difficulties in young children*. National Academy Press.

Suggate, S., Schaughency, E., McAnally, H., & Reese, E. (2018). From infancy to adolescence: The longitudinal links between vocabulary, early literacy skills, oral narrative, and reading comprehension. *Cognitive Development, 47*, 82–95. https://doi.org/10.1016/j.cogdev.2018.04.005.

Suggate, S. P. (2015). The Parable of the Sower and the long-term effects of early reading. *European Early Childhood Education Research Journal, 23*, 524–544. https://doi.org/10.1080/1350293X.2015.1087154.

Suggate, S. P., & Martzog, P. (2021). Preschool screen-media usage predicts mental imagery two years later. *Early Child Development and Care*, 1–14. https://doi.org/10.1080/03004430.2021.1924164.

Suggate, S. P. (2022, May 23). Data for Mental imagery skill predicts adults' reading performance. Retrieved from osf.io/knmp5.

Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

Wallace, C. E., & Russ, S. W. (2015). Pretend play, divergent thinking, and math achievement in girls: A longitudinal study. *Psychology of Aesthetics, Creativity, and the Arts, 9*(3), 296–305. https://doi.org/10.1037/a0039006

Wittgenstein, L. (1922). *Tractatus logico-philosophicus*. Kegan.

Zwaan, R. A. (1999). Embodied cognition, perceptual symbols, and situation models. *Discourse Processes, 28*(1), 81–88. https://doi.org/10.1080/01638539909545070

Zwaan, R. A., & Madden, C. J. (2004). Updating situation models. *Journal of Experimental Psychology Learning Memory and Cognition, 30*(1), 283–288. https://doi.org/10.1037/0278-7393.30.1.283. discussion 289-91.