Quotation accuracy in educational research articles

Ard W. Lazonder [a],[*], Noortje Janssen [b]

[a] *Behavioural Science Institute, Radboud University, the Netherlands*
[b] *School of Pedagogical and Educational Science, Radboud University, the Netherlands*

A R T I C L E   I N F O

A B S T R A C T

This study examined to what extent educational research articles are cited in accordance with content. Analysis of 500 randomly selected citations from articles published between 2016 and 2020 and listed in Web of Science Core Collection revealed an overall accuracy rate of 85%. Quotation accuracy was independent of a citation's specificity and bibliometric characteristics. However, the severity of quotation mistakes was associated with how specific the content of a referenced source is described and whether authors quote the work of a colleague or their own. These findings are discussed in light of established accuracy rates in other academic fields and measures to reduce quotation errors are proposed.

1. Introduction

The seed for writing this article was planted in 2016 when the first author published a meta-analysis in *Review of Educational Research*. This paper was well received and soon reached the status of highly-cited article in Web of Science Core Collection. Occasional citation checks revealed that several quotations were inaccurate: some contained minor inconsistencies while others twisted or even contradicted the study's purpose, results, or conclusions. The issue was brought up in a collegial conversation with the second author, who made a casual remark that sparked the idea to conduct a more systematic analysis of quotation accuracy in educational research articles.

As the opening anecdote implies, we use the term 'quotation accuracy' to refer to the standard practice of giving a complete and correct account of the content of a referenced study. Whether that source is accurately cited in the text (i.e., the in-text-citation contains the correct author-date information) and correctly provided in the reference list (i.e., reference components such as journal name, volume number and DOI link are accurate) was not the focus of our study. Nor did we address the use of previously published ideas or phrases, including one's own, without crediting the source (i.e., plagiarism).

Citing in accordance with content is a key pillar of scientific integrity. The educational research community has established professional values and practices that, among other things, serve to free scientific publications from bias, fabrication, and plagiarism. These integrity principles have been documented in ethical codes of conduct (e.g., AERA Code of Ethics, 2011) and are further specified in publication guidelines issued by professional associations such as the American Psychological Association (2020). These directions essentially appeal to an author's collegiality and responsibility: they should feel morally obligated to acknowledge the intellectual property of their peers and prevent that untruths and half-truths turn into conventional knowledge (Bareket et al., 2020; Rivkin, 2020).

To what extent educational researchers adhere to these recommendations has long been unknown but a recent quotation analysis in

---

the Learning Sciences established that the hallmark paper by Freeman et al. (2014) was misquoted in 26.01% of the assertions found in citing articles (Martella et al., 2021). Additional insights can be gleaned from bibliometric analyses in the medical sciences, where Fenton et al. (2000) examined the accuracy of 200 references from 4 otolaryngology-head and neck surgery journals. The results showed that 83% of the quotations were in accordance with content. Other examinations of medical journals found comparable accuracy rates, which ranged from 81.0 to 88.1% (Jergas & Baethge, 2015; Lukić et al., 2004; Mogull, 2017). Similar analyses of ecology journal articles yielded quotation accuracy rates of 85 and 65%, respectively (Teixeira et al., 2013; Todd et al., 2007). Quotation accuracy in multidisciplinary science journals (75%) falls in the same range (Smith & Cumberledge, 2020), while the rate in biomedical journals is somewhat higher (90.8%; Pavlovic et al., 2021). These findings illustrate that, although scholars quote each other's work rather truthfully, about one in five quotations contains a mistake.

So what can go wrong? In the least problematic case, a quotation contains a minor inconsistency or factual error "not severe enough to contradict a statement by citing authors" (Jergas & Baethge, 2015, p. 3). Suppose we would have stated in the previous paragraph that Fenton et al. (2000) analyzed 200 references from 4 *surgical* journals, then our assertion would be almost entirely correct due to one trivial mistake. A more serious error would be to claim that the observed accuracy rate was based on the analysis of 2000 instead of 200 references as this would give a wrong impression of the study's replicability. And in the worst case, our quotation would not reflect the content of the source material at all, for instance by claiming that Fenton et al. found an 83% correctness of bibliographic information in reference lists.

We postulate that the occurrence and seriousness of quotation errors depends on how detailed or specific the content of a referenced source is described. The odds of making a mistake in direct quotations like the one in the second sentence of the previous paragraph is negligible and any misquotation is probably minor because citing authors intend to give a verbatim reproduction of the source content. More and more serious errors are bound to occur when the content of another work is paraphrased or summarized as this requires authors to interpret and reword the source content. However, when authors refer loosely to a research article, for instance by making a broad statement about its research purpose or methods, the likelihood of incorrect quotations might be rather low.

What causes quotation errors is hard to grasp. Ignorance seems a highly implausible reason: scientists have all been taught to cite properly and pass this knowledge on to their students—and the internet is replete with how-to guides that convey the basic rules and rationale of crediting a source (e.g., Penders, 2018). Instead, we conjecture that most quotation errors are made unconsciously and, hence, without bad intentions. Such genuine mistakes may arise either from incorrect interpretations of the content of a referenced publication (Rivkin, 2020) or because the citing author took shortcuts. In the latter case, authors quote a familiar source from the top of their head without double-checking the paper, read new source material selectively—for example, only the abstract or main conclusion—or rely on secondary citations without checking their accuracy (Martella et al., 2021). These practices unfortunately are rather common among psychologists and educational scientists, no matter their years of experience (Klitzing et al., 2019).

However, because prior research has not determined quotation accuracy across the branches of educational science, it is important first to establish whether a quotation problem actually exists before trying to explain why. Correct quotations are essential to evidence-based decision making, a movement that has gained momentum in educational policy and practice (e.g., Tipton et al., 2021). If citing authors misquote their sources, erroneous claims may be propagated, accumulate into collective false beliefs (Bareket et al., 2020) and potentially adversely affect the professional practices of educators, curriculum designers, and policy makers. Quotation errors could also undermine the credibility of educational research (Harzing, 2002; Wilson, 2015) and cause the consumers of scientific knowledge as well as the general public to distrust the work of educational scientists (Jergas & Baethge, 2015). Such skepticism should be avoided in order to prevent educational policy makers and practitioners from neglecting the valuable scholarly insights that help improve the teaching and learning of future generations.

The present study therefore aimed to find out to what extent educational research articles are cited in accordance with content. Perfect accuracy should of course be the standard but the opening anecdote as well as our 30-plus years of experience in reading and reviewing scientific papers caused us to suspect that the actual accuracy level will be lower. Based on quotation analyses in other disciplines (Fenton et al., 2000; Jergas & Baethge, 2015; Lukić et al., 2004; Mogull, 2017; Teixeira et al., 2013) and the Martella et al. (2021) study in the Learning Sciences, the accuracy rate in education was expected to be in the range of 74–91%.

We theorized that the accuracy of a quotation depends on its specificity and this relationship was assumed to follow a U-shaped curve. Direct quotations represent the most specific type of quotation and their accuracy will probably be close to perfect for obvious reasons. Broad statements are at the other extreme of the specificity continuum. Due to their general nature, broad statements likely leave little room for quotation errors and were therefore expected to be only slightly less accurate than direct quotations. Paraphrases and summaries rank second and third in specificity, respectively. Both were expected to be less accurate than direct quotations and broad statements because they convey a reworded interpretation of the source content. As paraphrases contain more specific details than summaries, their quotation accuracy was predicted to be lower. Furthermore, the seriousness of quotation inaccuracies was assumed to show a similar pattern—that is, most errors in direct quotations and broad statements were expected to be minor whereas the more serious mistakes would prevail in paraphrases and summaries.

The study also examined whether and how quotation accuracy was related to an article's bibliometric characteristics. Five characteristics were included in this analysis: journal impact factor, article type, article length, self-citation, and country affiliation. These features were examined without explicit hypotheses because evidence from previous studies is typically mixed. For example, some observed more correct quotations in articles from high-impact journals (Haussmann et al., 2013; Pavlovic et al., 2021) whereas other studies found that quotation accuracy was independent of journal impact factor (Armstrong et al., 2018; Buijze et al., 2012). Similar inconsistent findings exist for self-citations (Haussmann et al., 2013; Pavlovic et al., 2021) and article length (Schulmeister, 1998; Todd et al., 2010), while previous analyses of article type incorporated writings typical of medical journals, such as clinical work, guidelines, and protocols (Buijze et al., 2012; Pavlovic et al., 2021) and may therefore not generalize to educational research articles.

Country affiliation was included because nonnative speakers of English might be at a disadvantage when understanding and quoting the source text (Gayle & Shimaoka, 2017; Man et al., 2004). However, both Haussmann et al. (2013) and Buijze et al. (2012) found no such effect.

2. Method

2.1. Literature search

We were interested in gathering a set of 25 highly-cited articles for this investigation. A highly-cited article was defined as an empirical study or systematic review with over 100 citations at the time of our literature search in August 2020. To obtain these 25 articles, we first conducted a literature search using the Social Science Citation Index (SSCI) database of the Web of Science Core Collection. The search was limited to the category 'Education and Educational Research' (461,047 entries) and further refined by document type ('article') and top papers ('highly cited in the field', i.e., more annual citations than 99% of the papers in social sciences). This query returned 392 publications. Secondly, we eliminated those articles that were cited less than 100 times, which reduced the number of highly-cited articles to 221. Thirdly, we randomly selected 25 articles from this set of 221 articles. The final sample of highly-cited articles is shown in Table 1.

A similar stepwise approach was used to compile a corpus of 500 citing articles. For each of the 25 highly-cited articles, we first searched for citing papers using the SSCI database. As our research aimed to portray the state-of-the-art in quotation accuracy, these cited reference searches were limited to the last five years (2016–2020). Table 1 shows how many citing articles were found in these searches. Next, we randomly selected 20 citing articles for each of the 25 highly-cited articles and verified that the corpus contained 500 unique articles. These samples represented between 5 and 20% of the total number of citing articles per highly-cited article ($M = 11.35$, $SD = 0.04$).

2.2. Coding of citing article features

Five bibliometric characteristics were extracted from the citing articles. The first characteristic, *impact factor* (IF), was used as a proxy for the relative importance of the journal where the citing article was published. Impact factors were obtained from the 2019 release of Clarivate's InCite Journal Citation Reports®. *Article type* indicated whether the citing article was a research article, review article or theoretical article. These categories were adapted from the typology described in the publication manual of the American Psychological Association (2020) to meet the specific purpose of this study. Research articles reported either original data from lab experiments, classroom studies, surveys and case studies, or a secondary analysis of existing data. Review articles synthesized empirical findings in the form of a meta-analysis, systematic review, or narrative review; commentaries discussing the papers in a special issue were also included in this category. Theoretical articles drew upon scholarly works to propose a new theory, modify an existing theory, or criticize widespread notions in the field. As most journals impose word limits, the characteristic *article length* reflected the space authors received to articulate and justify their thoughts. This opportunity was approximated by the citing article's

Table 1
Corpus of twenty-five highly-cited articles used in the study.

Author(s)	Year	Journal	Times cited
1. Alibali and Nathan	2012	Journal of the Learning Sciences	169
2. Bassok et al.	2016	AERA Open	130
3. Baumert et al.	2010	American Educational Research Journal	596
4. Bradshaw et al.	2014	Journal of School Health	104
5. Cheung and Vogel	2013	Computers & Education	249
6. Egalite et al.	2015	Economics of Education Review	104
7. Gershenson et al.	2016	Economics of Education Review	110
8. Girard et al.	2013	Journal of Computer Assisted Learning	258
9. Golonka et al.	2014	Computer Assisted Language Learning	216
10. Jacobs et al.	2010	Journal for Research in Mathematics Education	381
11. Jarodzka et al.	2010	Learning and Instruction	179
12. Jeynes	2012	Urban Education	194
13. Kuo et al.	2014	Internet and Higher Education	139
14. Lampert et al.	2013	Journal of Teacher Education	190
15. Li et al.	2012	Journal of Research in Reading	118
16. Manca and Ranieri	2013	Journal of Computer Assisted Learning	186
17. Means et al.	2013	Teachers College Record	202
18. Moll et al.	2014	Learning and Instruction	146
19. Nicol et al.	2014	Assessment and Evaluation in Higher Education	218
20. Ronfeldt et al.	2013	American Educational Research Journal	311
21. Spooren et al.	2013	Review of Educational Research	182
22. Tamim et al.	2011	Review of Educational Research	305
23. Van den Berghe et al.	2014	Physical Education and Sport Pedagogy	128
24. Wilder	2014	Educational Review	206
25. Wong et al.	2010	BMC Medical Education	200

number of pages, including the reference list and appendices. *Self-citation* indicated whether at least one of the authors also contributed to the highly-cited article. Self-citations were binary scored regardless of authorship order. Finally, the article's *country* affiliation was indicated by the corresponding address; articles from England, Scotland, Northern Ireland, and Wales were grouped under United Kingdom (UK).

2.3. Coding of quotations

Quotations were operationally defined as the passage referring to the designated highly-cited article from Table 1. This passage could be a phrase, a full sentence, or multiple sentences. One quotation was extracted from each citing article; in case the highly-cited article was referred to more than once, the first quotation was selected for use in the analyses. Each selected quotation was coded in terms of specificity and accuracy using the category levels in Table 2; Appendix A contains the number of accurate and inaccurate quotations by specificity level for each of the 25 highly-cited articles. Inter-rater agreement was assessed by having both authors code 50 randomly selected citing articles. They reached 96% agreement on the identification of quotation passages. Gwet's agreement coefficient (AC_1) of the specificity and accuracy codings was 0.68 and 0.80, respectively, which represents good inter-rater reliability (Gwet, 2014). Disagreements were resolved through discussion and did not point to any systematic difference in the authors' interpretation of the rubric. The first author coded the remaining quotations, consulting the second author when in doubt.

2.4. Data analysis

Quotation accuracy was descriptively analyzed by considering the total number of quotations in each of the four category levels. Congruence checks were performed to ensure that the quotations to each highly-cited article matched the overall distribution in scores. Accuracy scores were then dichotomized as either 'accurate' (i.e., the original category 'entirely correct') or 'inaccurate' (the remaining three categories combined); these binary scores were used in binomial tests to determine whether the overall accuracy rate was significantly less than perfect and if so, whether it fell in the predicted range of 78–91%.

Next, the four-point accuracy scores were analyzed against the specificity scores and the five bibliometric characteristics. The relationship between quotation accuracy and the continuous bibliometric characteristics (i.e., article length and impact factor) was examined by nonparametric Spearman's rho correlations. Loglinear analysis was performed to determine whether quotation accuracy was associated with the categorical variables. The likelihood ratio indicated that the final model was a good fit of the data, $\chi^2(561) = 45.96$, $p > .999$. Standardized residuals were used to break down the significant interactions and answer this study's research questions.

3. Results

3.1. Quotation accuracy

Our accuracy coding revealed that 425 of the 500 quotations were entirely correct, which boils down to an overall quotation accuracy rate of 85%. Of the remaining quotations, 15 contained an inconsequential mistake and were classified as 'almost correct' (3.0%); 22 quotations were considered party correct (4.4%) and 38 were incorrect (7.6%). Differences in quotation accuracy among the 25 highly-cited articles were minor ($SD = 0.10$, range 70–100%). Binomial tests indicated that, although the observed overall 85% accuracy rate was less than perfect, $p < .001$ (1-sided), it exceeded the expected lower bound of 74%, $p < .001$ (1-sided), but not the predicted 91% upper limit, $p < .001$ (1-sided). These results demonstrate that quotation accuracy in educational research articles is not perfect, but still well on par with accuracy rates found in other disciplines.[1]

Before continuing the quantitative analyses, it could be informative to illustrate what counted as an accurate or inaccurate quotation and why. The following quotation was entirely correct (the citing article identifier is indicated in brackets):

> … a previous study (Kuo et al., 2014), which showed that interaction between student and instructor and interaction between student and content in online learning environments were significant predictors of online students' course satisfaction, whereas interaction between student and student was not. [13.17]

Knowing that this assertion gives an accurate account of the Kuo et al. (2014) findings, the next one obviously does not.

> Social interactions are an essential element for the success of online learning (Kuo et al., 2014) … [13.10]

This quotation was coded as 'incorrect' because Kuo et al. (2014) examined what influences student satisfaction rather than the success of online learning, and marginalized the role of social interaction by concluding that "learner–learner interaction did

---

[1] 189 citing articles had more than one citation to the primary text. We sampled one of these subsequent quotations from each article to check for potential order effects. Results showed that 83.60% of the subsequent quotations were accurate. This estimate did not differ significantly from the 85% rate observed in the first quotations, $p = .324$. Accuracy of first and subsequent quotations was consistent in 76.19% of the cases and deviations were random, $\chi^2(1, N = 189) = 1.77$, $p = .183$.

Table 2
Categories used in the coding of quotations.

Dimension	Categories	Definition
Accuracy	1. Entirely correct	The quotation exactly reflects the content of the source material
	2. Almost correct	The quotation adequately reflects the content of the source material, but contains one trivial inconsistency
	3. Partly correct	The quotation reflects the content of the source material, but either contains one or more substantive inconsistencies or misses important nuances
	4. Incorrect	The quotation does not reflect the content of the source material
Specificity	1. Direct quotation	Verbatim reproduction of a passage from the source material, placed in quotation marks and/or with location information provided
	2. Paraphrase	Restatement, in the authors' own words, of specific details from the source material
	3. Summary	Condensed description of a conclusion or major idea from the source material
	4. Broad statement	General mentioning of the research goals, methods, or central theme of the source material

not appear to have any effect on students' satisfaction, and the effects of learner–instructor interaction were relatively weak" (p. 43).

Quotations in the category 'partly correct' contained information that either contradicts or could not be found in the highly-cited article. For example, Cheung and Vogel (2013) concluded in their highly-cited article that both attitudes and subjective norms positively predict behavioral intentions. One of the citing articles paraphrased this conclusion as "… individuals are more likely to conduct a specific behaviour if subjective norm and experiential attitude are strong" [5.08]. While this passage rightly links behavioral intentions to attitudes and social norms, the threshold raised in the second part of the sentence is inconsistent with Cheung and Vogel's findings. Other partly-correct quotations credited too much information to the highly-cited article. Below are two examples of this quotation error; the information that could not be found in the highly-cited article is italicized.

> Assessments of teaching quality and effectiveness can be beneficial for improving teaching quality, evaluating faculty performance, *and training future faculty* (Spooren et al., 2013). [21.02]

> More recently, Wilder (2013) suggested that parental expectations for their children could sway children's future achievements as parents' expectations were found to be strongly associated *with their own past educational experiences* (Wilder, 2013). [24.02]

Quotations were also considered 'partly correct' if they missed important nuances, which was the case in the following example:

> … Egalite et al. (2015) observed a 0.02 standard deviation greater gain in mathematics test scores and a 0.004 standard deviation greater gain in reading test scores for Black students matched to Black teachers in Florida classrooms relative to Black students with a White teacher, all else being equal. [6.14]

> This quotation is incomplete because Egalite et al. (2015) found greater reading gains in elementary grades only; assigning students to race-congruent teachers had no effect in the middle/high school grades. Furthermore, the 0.02 higher gain in mathematics could nowhere be found in Egalite et al. (2015). We suspect it might be a typo (0.02 should probably be 0.002—but that *SD* applied to middle/high school grades). So although this quotation reflects the general conclusion by Egalite et al., it does contain some consequential mistakes.

Quotations coded as 'almost correct' contained minor mistakes in, for example, the name of a survey instrument [4.12; 4.18], the subject-matter participants studied during the experiment [11.02; 20.15] or the type of research reported in the highly-cited article [24.13]. Two almost-correct quotations contained a minor substantive inconsistency: Wilder (2014) examined the effects of parental involvement (instead of maternal involvement; [24.14]) on academic achievement, and Bassok et al. (2016) argued that manipulatives (instead of blocks; [2.14]) are common materials in early childhood classrooms.

3.2. Characteristics associated with quotation accuracy

The next set of analyses determined whether the overall accuracy rate applied uniformly to all 500 quotations, and whether the severity of inaccuracies was associated with a quotation's specificity and bibliometric characteristics. Table 3 contains the descriptives of the categorical variables in the analysis; the continuous characteristics under investigation were article length ($M = 17.81$, $SD = 8.19$, range 3–52 pages) and journal impact factor ($M = 2.57$, $SD = 1.34$, range 0.49–7.85). The distribution of citing articles on both dimensions was positively skewed: 50% of the articles were between 12 and 22 pages long, and 50% of the articles were published in journals with an impact factor of 1.59–3.00.

Spearman correlations demonstrated that quotation accuracy depended neither on the number of article pages, $\rho = -0.07$, $p = .130$, nor on the journal's impact factor, $\rho = -0.04$, $p = .409$. Loglinear analysis of the characteristics in Table 3 produced a final model in which the two-way interactions were retained, $\chi^2(63) = 87.25$, $p < .001$. In order to answer this study's research questions, only interactions with quotation accuracy were considered further. Of these, the associations with specificity, $\chi^2(9) = 34.36$, $p < .001$, and

Table 3
Quotation accuracy by specificity and categorical bibliometric characteristics.

	Quotation accuracy				
	Entirely correct	Almost correct	Partly correct	Incorrect	Total
		Specificity			
Direct quotation	5	1	0	0	6
Paraphrase	103	3	10	4	120
Summary	197	3	11	30	241
Broad statement	120	8	1	4	133
Total	425	15	22	38	500
		Article type			
Research article	351	13	20	32	416
Review article	48	1	0	4	53
Theoretical article	26	1	2	2	31
Total	425	15	22	38	500
		Self-citation			
Yes	31	4	0	1	36
No	394	11	22	37	464
Total	425	15	22	38	500
		Country			
US	174	9	11	11	205
China	38	1	2	1	42
Australia	24	0	1	1	26
Germany	22	1	2	1	26
UK	23	0	0	1	24
Other	144	4	6	23	177
Total	425	15	22	38	500

self-citation, $\chi^2(3) = 8.15$, $p = .043$, proved significant. This means that quotation accuracy was somehow related to these two characteristics while being independent of the type of article, $\chi^2(6) = 5.68$, $p = .460$, and its county affiliation, $\chi^2(15) = 20.34$, $p = .159$.

Standardized residuals were used to break down the significant interactions. Regarding specificity, none of the standardized residuals for entirely correct quotations was significant, $-0.55 < z < 0.65$, $p > .513$. These outcomes signify that whether quotations were accurate or not, was independent of their specificity. Finer-grained investigation of the *inaccurate* quotations revealed that the number of almost-correct direct quotations tended to be higher than expected, $z = 1.93$, $p = .053$. For broad statements, the scores in this category were significantly higher than expected, $z = 2.01$, $p = .045$; broad statements were also less likely to be partly correct, $z = -2.00$, $p = .045$, and incorrect, $z = -1.92$, $p = .055$, than expected. Furthermore, more summaries than expected were incorrect, $z = 2.73$, $p = .001$, and more paraphrases than expected were partly correct, $z = 2.05$, $p = .040$.

Standardized residuals of the self-citation scores evidenced that the likelihood of making an entirely correct quotation was independent of whether authors cited the work of a colleague or their own, $z < 0.07$, $p > .942$. The scores in Table 3 further show that the inaccurate quotations were disproportionately distributed such that most self-citations were almost correct and more than half of the other-citations were incorrect. The former observation reached statistical significance, $z = 2.81$, $p = .005$, while the latter did not, $z = 0.29$, $p = .770$.

4. Discussion

Citing in accordance with content is a key element of scientific integrity in any academic discipline. This study estimated that quotation accuracy in educational research articles is 85%, with only 7.6% of the misquotations being entirely incorrect. These numbers are on par with quotation accuracy rates in medicine (Fenton et al., 2000; Jergas & Baethge, 2015; Lukić et al., 2004; Mogull, 2017) and ecology (Teixeira et al., 2013), and compare favorably to those in the Learning Sciences (Martella et al., 2021). Nonetheless, these mutual comparisons give little reason for optimism, if any, because 15% inaccuracy still means that, on average, one out of six to seven quotations contains a mistake.

The odds of a quotation being accurate did not depend on how specific the source content was referred to. This result contradicts our prediction that direct quotations and broad statements are accurate more often than summaries and paraphrases. The rather disproportionate distribution of specificity levels might have contributed to this nonsignificant outcome, in particular because the number of direct quotations was extremely low, so the one mistake we noticed caused an immediate 17% accuracy drop. The high proportions of correct paraphrases and summaries further suggest that the assumption underlying our hypothesis was incorrect or at least incomplete. We assumed that quotations would be less accurate to the extent that specific source content was reworded by citing authors. The study's findings suggest that the anticipated interpretation flaws are either rare or overcome, perhaps because authors who make a more specific quotation also pay more specific attention to the source. Our data lend indirect support to this conjecture because, by and large, quotation mistakes were less serious to the extent that assertions were more specific.

Quotation accuracy was independent of an article's bibliometric characteristics. These results imply that readers should not 'judge a book by its cover' and reject particular sources upfront for an assumed lack of quotation accuracy. Two bibliometric characteristics in

particular merit commenting. As most scholarly articles are written in English, it could be more difficult for nonnative authors to quote accurately. This was not the case, however, because quotation accuracy was independent of country affiliation. The Chinese and German scholars in our sample may of course have asked their peers for help or used the services of an English-language editing company, but language-focused copy editing usually does not target the correctness of quotations. So these exophonic writers are perhaps less fluent, but not less accurate than their native-English colleagues. The results further show that articles from high and low-impact journals have similar accuracy rates. This outcome seems at least partly due to our decision to select citing articles from Web of Science Core Collection—a database which applies high quality standards. Nonetheless, as the overall accuracy rate leaves room for improvement, journal editors and publishers aiming for both high impact and high quality might consider additional measures. Some suggestions are given below.

The severity of quotation mistakes was independent of all bibliometric characteristics except 'self-citation'. Although self-citations were accurate as often as references to someone else's work, the mistakes made in self-citations were less serious. Familiarity with one's own work is an obvious reason for this prevalence of minor mistakes, but it does not explain why 5 of the 36 self-citations were inaccurate in the first place. Perhaps authors quote their own work of the top of their head without checking the details, so minor mistakes are easily made, in particular when authors have many related publications or played a distant role in the cited research.

Given that quotation accuracy was largely independent of a citing article's bibliometric characteristics, it could be worthwhile to consider whether journal characteristics play a role. The journal impact factor did not, at least in this study, but other metrics might. One potentially distinguishing feature is the duration and quality of the peer review process. As thorough reviews take time and domain experts have many other professional commitments, the odds of detecting quotation mistakes may depend on the peer review timeframe. The number of referees, their credentials and the specificity of the journals' review guidelines might also contribute to quotation accuracy. Plagiarism checks, on the other hand, are less likely to raise quotation accuracy because the screening software reports *similarities* between a submitted manuscript and previously published sources. However, as most information on the peer review process is not publicly available, studies aspiring to investigate these characteristics should seek the help of editors and scientific publishers.

The frequency and severity of self-citation errors likely increases when authors rework their scientific publications into writings that serve a more applied purpose, such as professional journal articles and policy documents. These transcripts are typically short and urge authors to present their findings in bite-sized chunks, which comes at the loss of important nuances and, hence, increases the chance of what our rubric defined as 'partly correct' quotations. This conjecture was supported by Hui et al. (2020), who found a quotation accuracy rate of less than 60% in medical policy documents. So despite good intentions (practitioners and policy makers have insufficient knowledge and time to read scientific papers), the remedy can be worse than the disease. If this situation applies to public educational documents is unknown and should be examined in future research.

Future studies should also validate the present findings, which were based on the analysis of 500 references. Even though this corpus size compares well with studies in other fields, the collection was compiled based on a random sample of 20 articles that quoted one of 25 highly-cited papers randomly selected from Web of Science Core Collection. The included studies represent a mere 0.001% of the SSCI articles in the category 'Education and Educational Research'. Despite our systematic sampling method, we cannot guarantee that the estimated 85% accuracy rate generalizes to articles, journals and branches of educational science not included in our study. The present conclusions should therefore be interpreted with some caution, and extended replication studies are needed to consolidate our conclusions and differentiate accuracy rates across branches of educational science.

Future investigations should also broaden the scope of educational quotation analysis, for instance by examining the average number of accurate quotations in single articles or how key publications are cited (cf. Martella et al., 2021). Another possibility would be to include articles *not* indexed in Web of Science Core Collection or even go beyond journal articles by including conference papers and research reports. The latter resources in particular might be less accurate because their peer review guidelines tend to be more lenient. Similarly, different quotation accuracy rates could be found when references to other source materials are taken into account. Apparent options include highly-cited books, for instance the seminal volume 'Visible learning' (Hattie, 2009), monographs, and classic works such as the writings of John Dewey or Jean Piaget. As most of these publications are a long read and not readily accessible online, citing authors could be tempted to rely on secondary citations and, hence, unknowingly copy previous inaccuracies—a phenomenon known as the Chinese whispers' effect (Harzing, 2002). Incorporating these suggestions in future research will give a more complete picture of quotation accuracy in educational research papers.

Since the overall quotation accuracy rate is not optimal, we reiterate Jergas and Baethge's (2015) suggestions for improvements and add two proposals of our own. At the risk of stating the obvious, authors should be reminded to quote accurately, for instance through an awareness-raising statement in a journal's author guidelines and a tickbox in the online submission system to declare that all quotations have been checked for accuracy. To enforce these good intentions, editorial assistants could verify selected citations before a manuscript is send out for review. Journals could also ask reviewers to indicate in their comments whether quotes to known sources were accurate, or perhaps even add it as a separate scoring category to the online review form. Both options likely prevent misquotations because reviewers are invited based on their field expertise or because one of their papers is in the reference list. Once a manuscript is published online, the article's landing page could offer a feature to comment on inaccurate quotations; a tool that already offers this possibility is scite.ai. Finally, those involved in training the next generation of educational scientists should consider using quotation accuracy coding as a learning activity. It significantly enhanced our understanding of citation practices and we believe this experience is truly informative to master and PhD students.

In similar vein, our research could raise awareness of quotation accuracy among educational scientists. The results of the qualitative analysis exemplify that mistakes are easily made, and might cause scholars to attend more closely to a source when reading, reviewing and writing scientific papers. Although it is hard to foresee whether and to what extent this increased alertness will affect

research quality, drastic improvements seem unlikely because few of the observed misquotations were entirely incorrect. Nonetheless, we live in an era where misinformation and fake news are accelerating, so any self-respecting scientific discipline should avoid and act against unsupported assertions. The mere act of trying will send a positive sign to the scientific community and the general public.

To conclude, we estimated that 85% of the quotations in educational research articles is entirely correct. Quotation accuracy neither depends on how specific the content of a referenced source is described nor on the article's bibliometric characteristics. Quotation errors become less severe if authors refer to their own work, or paraphrase or summarize specific source content. We believe our study is an important first step in determining and understanding the possible existence of a quotation problem in educational science. Although the present findings are not immediately alarming, questions remain regarding the impact of inaccurate quotations on educational policy, teaching practices and the public opinion about the credibility of our academic field.

Author statement

Ard W. Lazonder: Conceptualization, Methodology, Validation, Formal analysis, Writing – original draft. Noortje Janssen: Conceptualization, Methodology, Validation, Formal analysis, Writing – review & editing.

Declaration of competing interest

None.

Appendix A

Number of accurate and inaccurate quotations by specificity level for each highly-cited article.

Article	Direct quotation		Paraphrase		Summary		Broad statement	
	Acc.	Inac.	Acc.	Inac.	Acc.	Inac.	Acc.	Inac.
1. Alibali and Nathan (2012)	0	1	4	1	7	0	6	1
2. Bassok et al. (2016)	0	0	6	1	9	3	1	0
3. Baumert et al. (2010)	1	0	3	0	8	1	7	0
4. Bradshaw et al. (2014)	1	0	1	1	2	2	10	3
5. Cheung and Vogel (2013)	0	0	5	2	4	1	8	0
6. Egalite et al. (2015)	0	0	7	1	10	1	1	0
7. Gershenson et al. (2016)	0	0	5	0	12	0	3	0
8. Girard et al. (2013)	0	0	1	1	6	1	10	1
9. Golonka et al. (2014)	0	0	4	1	9	0	6	0
10. Jacobs et al. (2010)	0	0	11	1	3	0	5	0
11. Jarodzka et al. (2010)	0	0	2	2	8	2	6	0
12. Jeynes (2012)	0	0	5	0	7	4	3	1
13. Kuo et al. (2014)	0	0	2	1	9	4	3	1
14. Lampert et al. (2013)	2	0	2	0	9	1	6	0
15. Li et al. (2012)	0	0	7	0	9	0	4	0
16. Manca and Ranieri (2013)	0	0	8	0	4	2	5	1
17. Means et al. (2013)	0	0	2	1	9	3	5	0
18. Moll et al. (2014)	0	0	2	0	13	2	3	0
19. Nicol et al. (2014)	0	0	4	0	11	1	4	0
20. Ronfeldt et al. (2013)	0	0	3	1	14	2	0	0
21. Spooren et al. (2013)	0	0	5	0	10	2	3	0
22. Tamim et al. (2011)	0	0	5	1	10	2	0	2
23. Van den Berghe et al. (2014)	0	0	2	0	3	6	9	0
24. Wilder (2014)	0	0	4	1	10	0	3	2
25. Wong et al. (2010)	1	0	3	1	2	3	9	1
Total	5	1	103	17	198	43	120	13

References

*References marked with an asterisk indicate the highly-cited articles used in the study.

AERA Code of Ethics: American Educational Research Association Approved by the AERA Council February 2011 (2011). Educational Researcher, 40(3), 145-156. https://doi.org/10.3102/0013189X11410403.

Alibali, M. W., & Nathan, M. J. (2012). Embodiment in mathematics teaching and learning: Evidence from learners' and teachers' gestures. *The Journal of the Learning Sciences, 21*(2), 247–286. https://doi.org/10.1080/10508406.2011.611446. *.

American Psychological Association. (2020). *Publication manual of the american psychological association* (7th ed.). American Psychological Association.

Armstrong, M. F., Conduff, J. H., 3rd, Fenton, J. E., & Coelho, D. H. (2018). Reference errors in otolaryngology–head and neck surgery literature. *Otolaryngology-Head and Neck Surgery, 159*(2), 249–253. https://doi.org/10.1177/0194599818772521

Bareket, R., Schonberg, M. A., & Schonmann, Y. (2020). Make quotations great again: A proposal to reduce false-knowledge. *BMJ Evidence-Based Medicine, 25*(1), 12–14. https://doi.org/10.1136/bmjebm-2018-111109

Bassok, D., Latham, S., & Rorem, A. (2016). Is kindergarten the new first grade? *AERA Open, 2*(1), 1–31. https://doi.org/10.1177/2332858415616358. *.

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., Klusmann, U., Krauss, S., Neubrand, M., & Tsai, Y.-M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal, 47*(1), 133–180. https://doi.org/10.3102/0002831209345157. *.

Bradshaw, C. P., Waasdorp, T. E., Debnam, K. J., & Johnson, S. L. (2014). Measuring school climate in high schools: A focus on safety, engagement, and the environment. *Journal of School Health, 84*(9), 593–604. https://doi.org/10.1111/josh.12186. *.

Buijze, G. A., Weening, A. A., Poolman, R. W., Bhandari, M., & Ring, D. (2012). Predictors of the accuracy of quotation of references in peer-reviewed orthopaedic literature in relation to publications on the scaphoid. *The Journal of Bone and Joint Surgery, 94-B*(2), 276–280. https://doi.org/10.1302/0301-620X.94B2.27618

Cheung, R., & Vogel, D. (2013). Predicting user acceptance of collaborative technologies: An extension of the technology acceptance model for e-learning. *Computers & Education, 63*, 160–175. https://doi.org/10.1016/j.compedu.2012.12.003. *.

Egalite, A. J., Kisida, B., & Winters, M. A. (2015). Representation in the classroom: The effect of own-race teachers on student achievement. *Economics of Education Review, 45*, 44–52. https://doi.org/10.1016/j.econedurev.2015.01.007. *.

Fenton, J. E., Brazier, H., De Souze, A., Huges, J. P., & McShane, D. P. (2000). The accuracy of citation and quotation in otolaryngology/head and neck surgery journals. *Clinical Otolaryngology and Allied Sciences, 25*(1), 40–44. https://doi.org/10.1046/j.1365-2273.2000.00322.x

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences of the United States of America, 111*(23), 8410–8415. https://doi.org/10.1073/pnas.1319030111

Gayle, A. A., & Shimaoka, M. (2017). Evaluating the lexico-grammatical differences in the writing of native and non-native speakers of English in peer-reviewed medical journals in the field of pediatric oncology: Creation of the genuine index scoring system. *PLoS One, 12*(2), Article e0172338. https://doi.org/10.1371/journal.pone.0172338

Gershenson, S., Holt, S. B., & Papageorge, N. W. (2016). Who believes in me? The effect of student-teacher demographic match on teacher expectations. *Economics of Education Review, 52*, 209–224. https://doi.org/10.1016/j.econedurev.2016.03.002. *.

Girard, C., Ecalle, J., & Magnan, A. (2013). Serious games as new educational tools: How effective are they? A meta-analysis of recent studies. *Journal of Computer Assisted Learning, 29*(3), 207–219. https://doi.org/10.1111/j.1365-2729.2012.00489.x. *.

Golonka, E. M., Bowles, A. R., Frank, V. M., Richardson, D. L., & Freynik, S. (2014). Technologies for foreign language learning: A review of technology types and their effectiveness. *Computer Assisted Language Learning, 27*(1), 70–105. https://doi.org/10.1080/09588221.2012.700315. *.

Gwet, K. L. (2014). Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among raters. In *Advanced analytics* (4th ed.).

Harzing, A. W. (2002). Are our referencing errors undermining our scholarship and credibility? *Journal of Organizational Behaviour, 23*(1), 127–148. https://doi.org/10.1002/job.125

Hattie, J. A. C. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement.* Routledge.

Haussmann, N. S., McIntyre, T., Bumby, A. J., & Loubser, M. J. (2013). Referencing practices in physical geography: How well do we cite what we write? *Progress in Physical Geography: Earth and Environment, 37*(4), 543–549. https://doi.org/10.1177/0309133313482135

Hui, A., Rains, L. S., Todd, A., Boaz, A., & Johnson, S. (2020). The accuracy and accessibility of cited evidence: A study examining mental health policy documents. *Social Psychiatry and Psychiatric Epidemiology, 55*(1), 111–121. https://doi.org/10.1007/s00127-019-01786-8

Jacobs, V. R., Lamb, L. L. C., & Philipp, R. A. (2010). Professional noticing of children's mathematical thinking. *Journal for Research in Mathematics Education, 41*(2), 169–202. https://doi.org/10.5951/jresematheduc.41.2.0169. *.

Jarodzka, H., Scheiter, K., Gerjets, P., & Van Gog, T. (2010). In the eyes of the beholder: How experts and novices interpret dynamic stimuli. *Learning and Instruction, 20*(2), 146–154. https://doi.org/10.1016/j.learninstruc.2009.02.019. *.

Jergas, H., & Baethge, C. (2015). Quotation accuracy in medical journal articles: A systematic review and meta-analysis. *PeerJ, 3*, e1364. https://doi.org/10.7717/peerj.1364

Jeynes, W. (2012). A meta-analysis of the efficacy of different types of parental involvement programs for urban students. *Urban Education, 47*(4), 706–742. https://doi.org/10.1177/0042085912445643. *.

Klitzing, N., Hoekstra, R., & Strijbos, J.-W. (2019). Literature practices: Processes leading up to a citation. *Journal of Documentation, 75*(1), 62–77. https://doi.org/10.1108/JD-03-2018-0047

Kuo, Y.-C., Walker, A. E., Schroder, K. E. E., & Belland, B. R. (2014). Interaction, internet self-efficacy, and self-regulated learning as predictors of student satisfaction in online education courses. *The Internet and Higher Education, 20*(1), 35–50. https://doi.org/10.1016/j.iheduc.2013.10.001. *.

Lampert, M., Franke, M. L., Kazemi, E., Ghousseini, H., Turrou, A. C., Beasley, H., Cunard, A., & Crowe, K. (2013). Keeping it complex: Using rehearsals to support novice teacher learning of ambitious teaching. *Journal of Teacher Education, 64*(3), 226–243. https://doi.org/10.1177/0022487112473837. *.

Li, H., Shu, H., McBride-Chang, C., Liu, H., & Peng, H. (2012). Chinese children's character recognition: Visuo-orthographic, phonological processing and morphological skills. *Journal of Research in Reading, 35*(3), 287–307. https://doi.org/10.1111/j.1467-9817.2010.01460.x. *.

Lukić, I. K., Lukić, A., Glunčić, V., Katavić, V., Vučenik, V., & Marušić, A. (2004). Citation and quotation accuracy in three anatomy journals. *Clinical Anatomy, 17*(7), 534–539. https://doi.org/10.1002/ca.10255

Manca, S., & Ranieri, M. (2013). Is it a tool suitable for learning? A critical review of the literature on facebook as a technology-enhanced learning environment. *Journal of Computer Assisted Learning, 29*(6), 487–504. https://doi.org/10.1111/jcal.12007. *.

Man, J. P., Weinkauf, J. G., Tsang, M., & Sin, D. D. (2004). Why do some countries publish more than others? An international comparison of research funding, English proficiency and publication output in highly ranked general medical journals. *European Journal of Epidemiology, 19*(8), 811–817. https://doi.org/10.1023/B:EJEP.0000036571.00320.b8

Martella, A. M., Yatcilla, J. K., Martella, R. C., Marchand-Martella, N. E., Ozen, Z., Karatas, T., Park, H. H., Simpson, A., & Karpicke, J. D. (2021). Quotation accuracy matters: An examination of how an influential meta-analysis on active learning has been cited. *Review of Educational Research, 91*(2), 272–308. https://doi.org/10.3102/0034654321991228

Means, B., Toyama, Y., Murphy, R. F., & Baki, M. (2013). The effectiveness of online and blended learning: A meta-analysis of the empirical literature. *Teachers College Record, 115*(3), 1–47. *.

Mogull, S. A. (2017). Accuracy of cited "facts" in medical research articles: A review of study methodology and recalculation of quotation error rate. *PLoS ONE, 12*(9), Article e0184727. https://doi.org/10.1371/journal.pone.0184727

Moll, K., Ramus, F., Bartling, J., Bruder, J., Kunze, S., Neuhoff, N., Streitau, S., Lyytinen, H., Leppanen, P. H. T., Lohvansuu, K., Tóth, D., Honbolygó, F., Csépe, V., Bogliotti, C., Lannuzzi, S., Démonet, J. F., Longeras, E., Valdois, S., George, F., & Landerl, K. (2014). Cognitive mechanisms underlying reading and spelling development in five European orthographies. *Learning and Instruction, 29*(1), 65–77. https://doi.org/10.1016/j.learninstruc.2013.09.003. *.

Nicol, D., Thomson, A., & Breslin, C. (2014). Rethinking feedback practices in higher education: A peer review perspective. *Assessment & Evaluation in Higher Education, 39*(1), 102–122. https://doi.org/10.1080/02602938.2013.795518. *.

Pavlovic, V., Weissgerber, T., Stanisavljevic, D., Pekmezovic, T., Milicevic, O., Lazovic, J. M., & Milic, N. (2021). How accurate are citations of frequently cited papers in biomedical literature? *Clinical Science, 135*(5), 671–681. https://doi.org/10.1042/CS20201573

Penders, B. (2018). Ten simple rules for responsible referencing. *PLoS Computational Biology, 14*(4), Article e1006036. https://doi.org/10.1371/journal.pcbi.1006036

Rivkin, A. (2020). Manuscript referencing errors and their impact on shaping current evidence. *American Journal of Pharmaceutical Education, 84*(7), ajpe7846. https://doi.org/10.5688/ajpe7846

Ronfeldt, M., Loeb, S., & Wyckoff, J. (2013). How teacher turnover harms student achievement. *American Educational Research Journal, 50*(1), 4–36. https://doi.org/10.3102/0002831212463813. *.

Schulmeister, L. (1998). Quotation and reference accuracy of three nursing journals. *Image-The Journal of Nursing Scholarship, 30*(2), 143–146. https://doi.org/10.1111/j.1547-5069.1998.tb01269.x

Smith, N., & Cumberledge, A. (2020). Quotation errors in general science journals. *Proceedings of the Royal Society A, 476*(2242), 20200538. https://doi.org/10.1098/rspa.2020.0538

Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching: The state of the art. *Review of Educational Research, 83*(4), 598–642. https://doi.org/10.3102/0034654313496870. *.

Tamim, R. M., Bernard, R. M., Borokhovski, E., Abrami, P. C., & Schmid, R. F. (2011). What forty years of research says about the impact of technology on learning: A second-order meta-analysis and validation study. *Review of Educational Research, 81*(1), 4–28. https://doi.org/10.3102/0034654310393361. *.

Teixeira, M. C., Thomaz, S. M., Michelan, T. S., Mormul, R. P., Meurer, T., Fasoli, J. V. B., & Silveira, M. J. (2013). Incorrect citations give unfair credit to review authors in ecology journals. *PLoS ONE, 8*(12), Article e81871. https://doi.org/10.1371/journal.pone.0081871

Tipton, E., Spybrook, J., Fitzgerald, K. G., Wang, Q., & Davidson, C. (2021). Toward a system of evidence for all: Current practices and future opportunities in 37 randomized trials. *Educational Researcher, 50*(3), 145–156. https://doi.org/10.3102/0013189X20960686

Todd, P. A., Guest, J. R., Lu, J., & Chou, L. M. (2010). One in four citations in marine biology papers is inappropriate. *Marine Ecology Progress Series, 408*, 299–303. https://doi.org/10.3354/meps08587

Todd, P. A., Yeo, D. C. J., Li, D., & Ladle, R. J. (2007). Citing practices in ecology: Can we believe our own words? *Oikos, 116*(9), 1559–1601. https://doi.org/10.1111/j.2007.0030-1299.15992.x

Van den Berghe, L., Vansteenkiste, M., Cardon, G., Kirk, D., & Haerens, L. (2014). Research on self-determination in physical education: Key findings and proposals for future research. *Physical Education and Sport Pedagogy, 19*(1), 97–121. https://doi.org/10.1080/17408989.2012.732563. *.

Wilder, S. (2014). Effects of parental involvement on academic achievement: A meta-synthesis. *Educational Review, 66*(3), 377–397. https://doi.org/10.1080/00131911.2013.780009. *.

Wilson, M. L. (2015). The persistence of misinformation. *American Journal of Clinical Pathology, 144*(3), 359–360. https://doi.org/10.1309/AJCPMBVI0SW4QHGU

Wong, G., Greenhalgh, T., & Pawson, R. (2010). Internet-based medical education: A realist review of what works, for whom and in what circumstances. *BMC Medical Education, 10*. https://doi.org/10.1186/1472-6920-10-12. Article 12, *.