Educ Psychol Rev (2013) 25:445–472
DOI 10.1007/s10648-013-9228-0

REVIEW ARTICLE

Designing Instructional Text in a Conversational Style: A Meta-analysis

Paul Ginns · Andrew J. Martin · Herbert W. Marsh

Published online: 7 May 2013

Abstract This article reviews research on the effects of conversational style on learning. Studies of conversational style have variously investigated "personalization" through changing instances of first-person address to second or third person, including sentences that directly address the learner; including more polite forms of address; and making the views and personality of the author more visible. Meta-analyses provided mixed support for a model of learning processes; statistically reliable average effects were found on self-reports of friendliness (d=0.46) and effective cognitive processing (d=0.62), but not learning assistance (d=0.16) and interest (d=0.15). Statistically reliable average effects on retention (d=0.30) and transfer (d=0.54) learning outcomes supported conversational-style redesigns across a range of potential moderators; the clearest apparent boundary condition for learning outcomes across the moderators under analysis was instructional time, with small, non-significant effects being found in studies longer than 35 min. Recommendations for future investigations are discussed.

Keywords Conversational style · Personalization · Instructional design · Meta-analysis

Expository text has been used for instruction for millennia, but students may often find such text dry, boring and lacking relevance to themselves. As a result, students may often engage superficially with the text and construct superficial or incorrect representations of the author's argument (for a discussion of students' perceptions of engagement with history textbooks, see Paxton 1997, 1999). Instructional writers and designers may draw on a range of methods to vitalise instructional text, but the justification for such methods may often be craft-based rather than theory-driven and empirically validated. Over the past two decades, educational researchers have experimentally tested a number of text amendments, such as text signals (Lorch 1989) or questions requiring elaborative interrogation (Menke and

P. Ginns (✉) · A. J. Martin
Faculty of Education and Social Work, The University of Sydney, Sydney, Australia
e-mail: paul.ginns@sydney.edu.au

H. W. Marsh
University of Western Sydney, Sydney, Australia

Pressley 1994), in the hope that relatively simple redesigns may generate substantial improvements in student engagement and learning. Using meta-analytic methods, the present study reviews research on the use of *conversational style* and its effects on learning processes and outcomes across a range of potential moderators.

Extant research has focused on several dimensions of conversational style. Of the studies reviewed here, the majority have aimed to personalize text through one or more stylistic changes. Studies of the *personalization effect* include changing text to emphasise first- or second-person forms of address rather than third-person forms. For example, a short (100-word) narration about the human respiratory system used by Mayer et al. (2004) replaced all instances of "the" with "your", e.g. "During inhaling, the [your] diaphragm moves down creating more space for the [your] lungs" (p. 391). In addition, some studies have also included sentences that directly address the learner; thus, the personalized version of instructions on lightning formation used by Moreno and Mayer (2000) began with the statement, "Let me tell you what happens when lightning forms. Suppose you are standing outside, feeling the warm rays of the sun heating up the earth's surface around you" (p. 732). Grouped with these studies, and discussed further below, are findings on *author visibility* (Paxton 2002) as this redesign both emphasises first-person writing and the revelation of personal beliefs and self, e.g. "But you don't have to trust me on this; Caesar's own point of view is spelled out in his book The Gallic Wars…" (p. 244). This meta-analysis also contrasts personalization studies with a smaller number of *politeness* studies, which examined learning from directly worded text (e.g. "Save the factory now") with more "face-saving" text (e.g. "Why don't we save the factory now?"; Wang et al. 2008, p. 104).

Mayer (2005a, b) argued that the presence of social cues in human–computer interaction, such as a conversational style, will activate sense-making efforts of the same kind that occur during social (human–human) interactions. Such efforts are based on a listener's assumptions that a speaker is trying to "make sense"; hence, there is a strong implicit drive on the part of the listener to cooperate in the sense-making activity (Grice 1975). Mayer argues that this activity results in deeper cognitive processing and, as a result, better learning outcomes. Figure 1 illustrates this chain of reasoning as well as including potential moderators (discussed below) of conversational-style redesigns.

To a greater or lesser extent, each of the above redesigns aim to generate instructional text that mimics the conventions of conversation—that is, communication between people rather

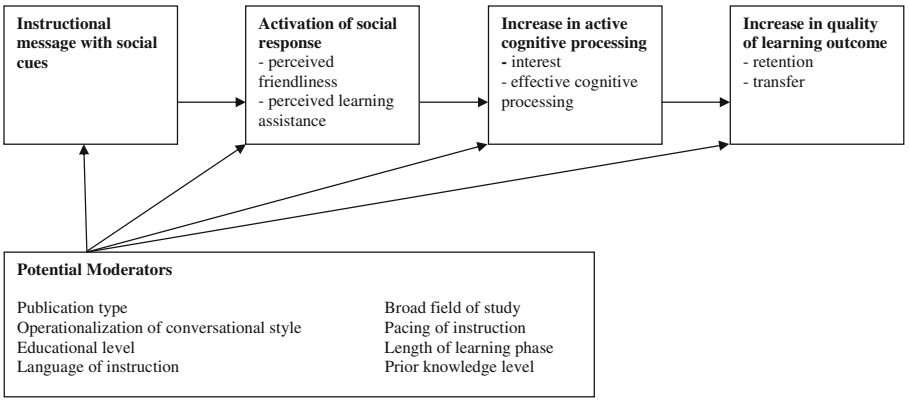

Fig. 1 Hypothesised effects of conversational style on learning processes and outcome measures, and potential moderators

than just between a learner and an "inert" text. Theoretical underpinnings of this body of research are first discussed, followed by a review of the extant research as the basis for potential moderators.

The Role of the Self and Social Cues in Learning

There has been long-standing recognition of the role of the self in learning, particularly by theorists working within humanistic or cognitive paradigms. From a cognitive perspective, there are several substantial bodies of research linking the self, cognitive processing and learning. Spiro (1977) considered a number of circumstances in which novel information from text might be integrated into existing schemata, as opposed to being read and understood in isolation. Consistent with the research described above, he identified interests and attitudes as playing a key role, but also argued that the presence of speaker–hearer and writer–reader relationships would activate cognitive processing, as would the explicit use of certain linguistic cues: "for example, an author explicitly informing the reader that some point under discussion will be illuminated from a specified, broader perspective" (p. 157). Lastly, he argued that kinds of discourse may vary substantially in the degree to which they can be easily related to an individual's existing schemata.

Drawing on Spiro's (1977) arguments, Bretzing and Kulhavy (1981) investigated the impact of the note-taking method and text formality on recall from a text on traffic accidents. Bretzing and Kulhavy defined formality as the extent to which to-be-learned information can be comprehended by all or most readers: thus, "a low-formality passage would be designed for communication with learners who were well acquainted with the terminology and examples used in the text [while] a highly formal passage is written to maximize communication with as few personally identified instances as possible" (p. 242). An excerpt of the low-formality passage is as follows:

> Driving too fast contributes to about one third of all accidents you may be involved in. The faster you are going when you crash, the more chance you have of being hurt or killed. You have less control over the car at high speeds. The new speed limit of 55 mph has probably reduced your chances of dying in a car accident.

The corresponding high-formality text was as follows:

> In 1974, "speed too fast" was reported as a contributory factor in 31 % of all fatal traffic accidents. High impact results in more severe damage to vehicles. The 55 mph maximum speed limit was responsible for most of a recent reduction in traffic fatalities.

The passages were otherwise matched on length and readability. Undergraduate students who read the low-formality version of the text passage recalled significantly more idea units from the text than those who studied the high-formality version. Interpreting these results, Bretzing and Kulhavy (1981) suggested that "low-formality prose is better suited for updating current schemata, thus improving recall over a highly formal presentation" (p. 248).

Narratives and Comprehension

Moreno and Mayer (2000) noted that the potential benefits of personalized messages may also be understood through research on narrative comprehension. In contrast to other genres

such as definition and exposition, text written in a narrative genre tends to have more concrete features (e.g. characters, objects, plot) than other genres and, like low-formality text, can be understood more easily in light of everyday experience (Graesser et al. 2002; Graesser and Ottati 1996; Haberlandt and Graesser 1985). Summarising results from Graesser et al. (1980a, b), Graesser et al. (2002) asserted that "narrative text is recalled approximately twice as well as expository text and is read approximately twice as fast" (p. 240), but also noted that the precise reasons for these effects are unclear. The possibilities raised include the correspondence of narrative text with everyday experiences; the similarity of oral conversation to narrative text; the generation of more vivid mental images; or narratives being more interesting, and hence more motivating, to read. However, Graesser et al. cautioned that this last hypothesis has not been supported in studies using interestingness ratings (e.g. Graesser et al. 1980a, b).

The Visible Author

Another related line of research has investigated the effects of a "visible author" in educational texts (Nolen 1995; Paxton 1997, 2002). Nolen (1995) contrasted visible authors with anonymous authors, who "remain invisible to readers through the use of rhetorical devices, including passive construction, use of the third rather than the first person, and by not directly discussing their personal history, views, or personality" (p. 47). Paxton (2002) investigated the effect of a visible author in a reading-to-write task on Julius Caesar with high school history students. Compared to students in the anonymous author condition, students in the visible author condition wrote substantially longer essays, a result Paxton argued as significant since essay length is often substantially correlated with other measures of writing quality (Bereiter and Scardamalia 1987). However, using independent ratings, the visible author condition outperformed the anonymous author condition on only two criteria (use of texts and personal agency), with no statistically significant differences on the remaining criteria (macrostructures, position taking and use of multiple perspectives).

The Personalization Principle

Mayer's cognitive theory of multimedia learning (CTML; Mayer 2008, 2009) has generated a substantial evidence base to guide instructional design based on the core assumptions of dual-processing channels (visual and auditory), a limited capacity working memory and the benefits of active learning. Whilst the focus of most CTML-based instructional designs has been on cognitive considerations, several have additionally drawn on theories of narrative and social interaction to generate learning process and outcome hypotheses. Theory and research described above informed initial theorizing for CTML's personalization principle: "people learn better from multimedia presentations when words are in conversational style rather than formal style" (Mayer 2009, p. 242). In addition, and reflecting CTML's focus on design for computer-based instruction, initial theorizing also drew on Reeves and Nass's (1996) findings that people are predisposed to apply the same dynamics of human–human interactions to human–computer interactions. Incorporating social cues into human–computer interaction, such as through conversational style, is held to generate identical sense-making efforts to those occurring during human–human interactions. Mayer posits that this activity leads to deeper cognitive processing as well as higher quality learning. In studies investigating the personalization principle, instructional texts were rewritten to include additional sentences which directly addressed the learner (e.g. "Let me tell you what happens when lightning forms"). In addition, sentences were rewritten in a conversational style by

using first- and second-person sentence constructions (you, I and your) rather than third-person constructions.

Moreno and Mayer's (2000) initial study examining personalization through conversational style involved college students participating in one of five experiments. Experiments 1 and 2 aimed to determine whether a personalization effect could be established using a simple multimedia-learning task followed by a retention and problem-solving transfer test. Experiment 2 replicated experiment 1; however, narrated speech was replaced with on-screen text. In experiment 3, participants played a multimedia game centred on environmental concepts, with the assistance of a pedagogical agent's speech in either personalized or non-personalized language. Experiment 4 was a replication of experiment 3, with the narration by a pedagogical agent being replaced by on-screen text. Experiment 5 also replicated experiment 3, except that the rhetorical questions asked during instruction in experiment 3 were excluded. Across the five experiments, students who learned by means of a personalized explanation were better able to transfer their knowledge to problem-solving tasks than students who received a non-personalized explanation.

A subsequent study by Moreno and Mayer (2004) sought to investigate the effects of personalization in conjunction with immersion. Immersion aims to create a sense of physical presence in learning, with the immersion hypothesis stating that "a greater level of immersion should lead learners to experience a greater sense of physical presence in which he or she has a feeling of being within the game environment" (Moreno and Mayer 2004, p. 166), with the expectation that participants immersed in a learning experience will achieve deeper learning, experience less difficulty, score higher on tests of retention and transfer, and rate the program as more friendly, helpful and easy to use. The latter predictions were tested via a survey asking participants to rate the program's levels of friendliness, helpfulness, difficulty, motivation and interest on a ten-point scale. The study found that participants in the personalized groups rated the program as being more friendly and helpful, and less difficult, than did students in the non-personalized groups whilst achieving higher scores on tests of retention and transfer.

Mayer et al. (2004) used multimedia instructional materials on the function of the human respiratory system accompanied by a retention test, a transfer test and an interest rating survey to test personalization's potential for increasing learner interest, thus making available cognitive capacity for the processing of new information (p. 389). Similar to Moreno and Mayer's study (2004), their "modest and focused" (p. 389) approach to personalization involved changing "the" to "your" in 12 places throughout the narrated animation. Across three studies, participants in the personalized group achieved better results on the transfer test than the non-personalized group, but both groups achieved relatively similar results on the retention test. Mayer and colleagues interpreted these differential results across retention and transfer as evidence of the potential of personalization to encourage more extensive active cognitive processing. Such processing is necessary, they argued, if a deeper understanding of the topic—supporting problem-solving transfer—is to be attained. In contrast, they argued that the lower-level outcome of retention will not index the effects of a deeper understanding as reliably as transfer.

The Politeness Effect

Recent research on the effects of conversational style has returned to Reeves and Nass's (1996) studies of people's social responses to information and communication technologies. Wang et al. (2008) argued that, if Reeves and Nass's arguments are correct, the effectiveness of animated pedagogical agents should depend on the extent to which their behaviour

accords with social norms. According to the politeness theory (Brown and Levinson 1987), across all cultures, people wish to maintain positive face by appearing desirable to others whilst also wishing to maintain negative face by retaining autonomy. Certain communicative acts may threaten positive and/or negative face, although the impact of such acts may depend on several factors, such as cultural background and power distance between the speaker and listener. Using politeness strategies can reduce both positive and negative face threats. An initial study by Mayer et al. (2006) found that, consistent with politeness theory, learners asked to imagine interacting with an on-screen pedagogical agent rated direct commands (e.g. "Click the ENTER button") as low on both negative and positive politeness; guarded suggestions (e.g. "You may want to click the ENTER button") and questions (e.g. "Do you want to click the ENTER button?") as highest in negative politeness; and statements of common purpose (e.g. "Let's click the ENTER button") as highest in positive politeness.

The above results informed the design of polite pedagogical agents in a subsequent series of studies. Wang et al. (2008) found that students who received polite feedback from a pedagogical agent about their progress through a computer-based factory simulation learned more than those who received direct feedback; however, there were no statistically reliable differences between conditions on ratings of self-efficacy, sense of control, interest or tutor helpfulness. Other studies of politeness effects have found mixed results. McLaren et al. (2007) conducted a classroom-based experiment in which high school students received either polite or direct problem statements, hints and error messages from a computer-based stoichiometry tutorial, but found no statistically reliable difference between conditions on the proportion of correct steps taken during the tutorial. An overall politeness effect was also not found in a larger ($N = 132$) classroom-based experiment with similar materials (McLaren et al. 2011a), although post hoc analyses suggested a politeness effect for those students who made the most errors during tutoring.

Replications, Extensions and Limitations

Reviewing the theoretical basis for the personalization of instructional material, Mayer et al. (2004) argued "using the self as a reference point increases the learner's interest, which in turn encourages the learner to use available cognitive capacity for active cognitive processing of the incoming information during learning. The deeper processing results in the more meaningful learning as indicated by better transfer test performance" (p. 391). However, in the series of experiments of Mayer et al. (2004), supporting evidence for these specific hypotheses was not forthcoming. Ratings of interest by participants in the personalized condition were not statistically different from those in the non-personalized condition. Similarly, ratings of difficulty of the materials did not differ reliably between conditions. The authors noted the possibility that null results on these measures could be due to defective instruments. In discussing limitations of their series of experiments, Mayer et al. noted, "…the missing links in our theoretical account…concern measures of interest (or personal relevance) and measures of depth of processing during learning…either our measurement instruments are somewhat defective or our hypothesis needs revision. Therefore, more focused research aimed at measuring interest and cognitive engagement would be helpful" (p. 394). Alternatively, these results might be the product of a sampling error, and a meta-analysis of all available results would provide a more powerful test of these hypotheses.

Research by Mayer and colleagues on the personalization principle has generated a number of attempts to replicate and extend the above findings. For example, one line of investigation has concerned the extent to which personalization effects can be

generated in languages other than English, given that languages can differ substantially in the ways in which formality is signalled. Kartal (2007), using Turkish as the language of instruction, did not find any statistically reliable differences in programme ratings (e.g. interest, perceived difficulty) or learning outcomes between a non-personal formal style condition and a personalized formal style condition. However, in a follow-up study, Kartal (2010) added an additional condition (personalized informal) and found that participants in both the personalized-formal and personalized-informal conditions rated materials as more interesting, more friendly, more motivating and less difficult (contrary to the results of Mayer et al. 2004, described above) than those in the non-personalized condition. Participants in the personalized-informal condition also outperformed those in the non-personalized condition on both retention and transfer tests, but the differences between the personalized-formal and the non-personalized conditions on retention and transfer were not statistically reliable. Other attempts to replicate initial findings in other languages have not been successful; for example, using instructional materials written in Flemish, Clarebout and Elen (2007) found that students who learned with the assistance of an on-screen pedagogical agent learned more effectively when the advice was written in third person rather than first or second person.

A number of researchers have also tested the personalization principle across longer acquisition periods, and in field rather than laboratory studies, but the results have been mixed. McLaren et al. (2006) failed to find an effect of incorporating a conversational style into the instruction, feedback and hints provided by a web-based intelligent tutoring system on stochiometry. Likewise, in a large study including over 600 students, Yeung et al. (2009) did not find a reliable personalization effect on university chemistry students' learning from online pre-laboratory modules. These results call into question the efficacy of personalization in more realistic educational settings and suggest that the effect may not always be robust when scaled up beyond the few minutes of study typical of laboratory studies.

Research on the expertise-reversal effect (cf. Kalyuga 2007), following earlier aptitude–treatment interaction research focusing on prior knowledge, has prompted some researchers to investigate whether the effects of conversational style are moderated by expertise levels. The study of politeness effects of Wang et al. (2008) within a virtual factory found that the positive effects of politeness were more pronounced for participants without an engineering background. Using a median split on a test of prior knowledge, Stiller and Jedlicka (2010) found that personalization of instructional text led to high school students with lower prior knowledge levels performing at higher levels on both recall and transfer tasks. In contrast, for students with higher prior knowledge levels, no significant difference was found between conditions on the recall test, but on the transfer test, students who studied non-personalized materials outperformed those who studied personalized materials. Lastly, based on a median split of students' self-reports of chemistry prior knowledge, McLaren et al. (2011b) also found an expertise-reversal effect. College students with lower prior knowledge who studied a polite computer-based stoichiometry tutorial performed better on both immediate and delayed problem-solving tests. In contrast, for students with higher prior knowledge, whilst the mean scores on both immediate and delayed tests favoured those who studied non-polite materials, these differences were not statistically reliable.

Previous reviews of conversational-style redesigns have focused on studies conducted by Mayer and colleagues (e.g. Mayer 2001, 2005a, b, 2009). In each of

these reviews, Mayer argued that *transfer* should be the focal dependent variable under investigation, reflecting a contemporary view of instruction that supports not only retention of core elements of a lesson but also the capacity to use what has been learned to solve novel problems. In the most recent review, Mayer (2009) reported a median effect on transfer test performance across 11 experiments of 1.11. The present meta-analysis builds on this review by meta-analysing effects on retention as well as the core outcome of transfer; by meta-analysing available learning process results as well as outcomes; and by including studies beyond those conducted by Mayer and colleagues.

Hypotheses

The research reviewed above suggests that a number of relatively minor adjustments to instructional text—including the use of first and second rather than third person, directly addressing the reader, revealing personal beliefs and/or using polite forms of address—may have substantial impacts on learning outcomes. However, the variability in results across studies raises important questions about the generalizability of such effects. In addition, hypotheses regarding the underlying reasons for these effects have been supported in some studies, but not in others; in many cases, null findings could potentially be attributed to insufficient experimental power to detect smaller effects. Meta-analysis of both learning process and outcome findings can support more powerful estimates of population parameters from the available body of research through the statistical aggregation of experimental findings.

The following major hypotheses will be tested, following Mayer's (2009) model of the process by which conversational style affects learning processes and outcomes (see Fig. 1).

Learning Processes

1. Conversational instructional text generates a more social response to learning materials, as indexed by *perceived friendliness* (e.g. "How friendly was the computer that you interacted with?"; Mayer and Moreno 2004, p. 168), compared to formal instructional text.
2. Conversational instructional text generates a more social response to learning materials, as indexed by *perceived learning assistance* (e.g. "How helpful is this material in learning about plant design)?"; Mayer and Moreno 2004, p. 168), compared to formal instructional text.
3. Conversational instructional text generates higher levels of *interest* (e.g. "How interesting was this lesson?"; Mayer et al. 2004, p. 393) compared to formal instructional text.
4. Conversational text supports more *effective cognitive processing* (e.g. "How difficult was it for you to learn this lesson?"; Mayer et al. 2004, p. 393), as indexed by lower ratings of difficulty during the learning phase.

Learning Outcomes

5. Conversational instructional text leads to enhanced learning compared to formal instructional text, as measured by tests of *retention* (e.g. "Write an explanation of how the respiratory system works"; Mayer et al. 2004, p. 391).

6. Conversational instructional text leads to enhanced learning compared to formal instructional text, as measured by tests of *transfer* (e.g. "Suppose you are a scientist trying to improve the human respiratory system. How could you get more oxygen into the bloodstream faster?"; Mayer et al. 2004, p. 391).

Method

Literature Search

The literature search was conducted by searching PsychINFO, ERIC and Google Scholar for the keywords personalization, Personalization, personalization principle, Personalization principle, conversational style, self-reference, self-referencing, self reference or self referencing, up to August 2012; searches of studies citing earlier studies using Science Citation Index; and examination of reference lists of individual articles.

Inclusion Criteria

The inclusion criteria for the meta-analysis were as follows. Studies were included if (a) the language of the report was English; (b) the study was based on an experimental or quasi-experimental design; (c) the independent variable of conversational style was not confounded with other variables (e.g. Inglese et al. 2007); (d) a statistic from which a standardised mean difference (*d*) effect size could be computed was available (e.g. means and standard deviations; *t* or *F* statistics); and (e) results for at least one learning outcome were reported.

It should be noted that whilst the search terms described above include self-reference, our purpose in this study was not to replicate Symons and Johnson's (1997) meta-analysis, which focused on aggregating results from laboratory studies using simple recognition or list recall tasks. Rather, the present study focuses on higher-order learning outcomes (Anderson et al. 2001) from substantially more complex learning materials than typical self-reference effect studies; the inclusion of self-reference in the search terms was done in order to capture any studies with a focus on conversational style. Only one such study (D'Ailly et al. 1995) was identified, and this study was not included in the present meta-analyses as it used reaction time as a dependent variable rather than a test of learning.

Based on the above criteria, 74 effects, representing effects on the learning outcomes and/or processes of 3,312 students, were coded from results given in 16 journal articles, 4 conference papers, and 2 PhD or Honours dissertations. These effects are presented in Table 1. The standardised mean difference was defined as the difference between the means of the conversational style and non-conversational style conditions divided by the pooled standard deviation, corrected for the slight bias associated with small sample sizes (Hedges and Olkin 1985).

In many cases, investigations of the effects of conversational style investigated effects across multiple learning processes and outcomes. The approach to meta-analysis employed here assumes that all comparisons across the same construct (e.g. interest, retention) are independent. For the most part, only one measure per construct was reported in each study. Where more than one measure was reported, effect sizes were calculated by averaging across measures; for example, McLaren et al. (2011a, b) reported results across both immediate and delayed post-tests of learning, and effects were calculated from the average of these sets of results. Results for similar constructs were also averaged; thus, Moreno and Mayer (2004)

Table 1 Summary of results included for meta-analysis

Study	Variate type	Standardised mean difference (*d*)	Focus	Educational level	Language	Experimental material format	Broad field of study	Pacing of instruction	Length of learning phase (min)	Prior knowledge level
Bretzing and Kulhavy (1981)	Retention	0.67	Personalization	College	English	Paper	Social science	System-paced	15	Mixed
Wagner et al. (1998)	Friendliness	0.54	Personalization	College	English	Paper	Social science	Self-paced	30	Mixed
	Learning assistance	0.44	Personalization	College	English	Paper	Social science	Self-paced	30	Mixed
	Retention	0.42	Personalization	College	English	Paper	Social science	Self-paced	30	Mixed
Moreno and Mayer (2000), Exp. 1	Retention	0.15	Personalization	College	English	Computer	Science	Self-paced	15	Low
	Transfer	1.02	Personalization	College	English	Computer	Science	Self-paced	15	Low
Moreno and Mayer (2000), Exp.2	Retention	0.20	Personalization	College	English	Computer	Science	Self-paced	15	Low
	Transfer	1.50	Personalization	College	English	Computer	Science	Self-paced	15	Low
Moreno and Mayer (2000), Exp. 3	Retention	0.96	Personalization	College	English	Computer	Science	Self-paced	26	Low
	Transfer	1.80	Personalization	College	English	Computer	Science	Self-paced	26	Low
Moreno and Mayer (2000), Exp. 4	Retention	0.62	Personalization	College	English	Computer	Science	Self-paced	26	Low
	Transfer	1.46	Personalization	College	English	Computer	Science	Self-paced	26	Low
Moreno and Mayer (2000), Exp. 5	Retention	0.61	Personalization	College	English	Computer	Science	Self-paced	26	Low
	Transfer	1.06	Personalization	College	English	Computer	Science	Self-paced	26	Low

Table 1 (continued)

Study	Variate type	Standardised mean difference (*d*)	Focus	Educational level	Language	Experimental material format	Broad field of study	Pacing of instruction	Length of learning phase (min)	Prior knowledge level
Paxton (2002)	Transfer	0.50	Visible author	Junior high school	English	Paper	Social Science	Self-paced	Not stated	Mixed
Moreno and Mayer (2004)	Retention	0.56	Personalization	College	English	Computer	Science	Self-paced	15	Low
	Transfer	2.33	Personalization	College	English	Computer	Science	Self-paced	15	Low
	Friendliness	0.82	Personalization	College	English	Computer	Science	Self-paced	15	Low
	Effective processing	0.59	Personalization	College	English	Computer	Science	Self-paced	15	Low
Mayer et al. (2004), Exp. 1	Retention	0	Personalization	College	English	Computer	Science	System-paced	1	Mixed
	Transfer	0.52	Personalization	College	English	Computer	Science	System-paced	1	Mixed
Mayer et al. (2004), Exp. 2	Retention	−0.23	Personalization	College	English	Computer	Science	System-paced	1	Mixed
	Transfer	0.96	Personalization	College	English	Computer	Science	System-paced	1	Mixed
Mayer et al. (2004), Exp. 3	Interest	0.51	Personalization	College	English	Computer	Science	System-paced	1	Mixed
	Effective processing	0.08	Personalization	College	English	Computer	Science	System-paced	1	Mixed
	Retention	0.38	Personalization	College	English	Computer	Science	System-paced	1	Mixed
	Transfer	0.77	Personalization	College	English	Computer	Science	System-paced	1	Mixed
Dunsworth (2005)	Learning assistance	0.19	Personalization	College	English	Computer	Science	Self-paced	14	Low
	Retention	0	Personalization	College	English	Computer	Science	Self-paced	14	Low
	Transfer	0.37	Personalization	College	English	Computer	Science	Self-paced	14	Low
	Learning assistance	0.19	Personalization	College	English	Computer	Science	Self-paced	14	Low
McLaren et al. (2006)	Retention	−0.13	Personalization	College	English	Computer	Science	Self-paced	60	Mixed

Table 1 (continued)

Study	Variate type	Standardised mean difference (*d*)	Focus	Educational level	Language	Experimental material format	Broad field of study	Pacing of instruction	Length of learning phase (min)	Prior knowledge level
Kartal (2007)	Friendliness	0.23	Personalization	College	Turkish	Computer	Science	Self-paced	Not stated	Mixed
	Learning assistance	0.08	Personalization	College	Turkish	Computer	Science	Self-paced	Not stated	Mixed
	Interest	−0.34	Personalization	College	Turkish	Computer	Science	Self-paced	Not stated	Mixed
	Effective processing	0.09	Personalization	College	Turkish	Computer	Science	Self-paced	Not stated	Mixed
	Retention	0.41	Personalization	College	Turkish	Computer	Science	Self-paced	Not stated	Mixed
	Transfer	−0.04	Personalization	College	Turkish	Computer	Science	Self-paced	Not stated	Mixed
Clarebout and Elen (2007)	Transfer	−0.52	Personalization	Junior high school	Flemish	Computer	Industrial design	Not stated	50	Low
Robertson (2008)	Interest	0.21	Personalization	College	English	Paper	Science	System-paced	25	Mixed
	Retention	0.32	Personalization	College	English	Paper	Science	System-paced	25	Mixed
Wang and Johnson (2008)	Retention	0.14	Politeness	Mixed levels of adult education	English	Computer	Language arts	System-paced	120	Not stated
Wang et al. (2008)	Learning assistance	−0.76	Politeness	College	English	Computer	Engineering	Not stated	35	Mixed
	Interest	0.05	Politeness	College	English	Computer	Engineering	Not stated	35	Mixed
	Retention	0.69	Politeness	College	English	Computer	Engineering	Not stated	35	Mixed
Yeung et al. (2009), Exp. 1	Retention	0.10	Personalization	College	English	Computer	Science	Self-paced	Not stated	Not stated
Yeung et al. (2009), Exp. 2	Retention	0.32	Personalization	College	English	Computer	Science	Self-paced	Not stated	Low

Table 1 (continued)

Study	Variate type	Standardised mean difference (*d*)	Focus	Educational level	Language	Experimental material format	Broad field of study	Pacing of instruction	Length of learning phase (min)	Prior knowledge level
Yeung et al. (2009), Exp. 3	Retention	0.15	Personalization	College	English	Computer	Science	Self-paced	Not stated	Medium
Yeung et al. (2009), Exp. 4	Retention	0.50	Personalization	College	English	Computer	Science	Self-paced	Not stated	High
Yeung et al. (2009), Exp. 5	Retention	0.28	Personalization	College	English	Computer	Science	Self-paced	Not stated	High
	Transfer	0.27	Personalization	College	English	Computer	Science	Self-paced	Not stated	High
Son and Goldstone (2009)	Retention	−0.30	Personalization	College	English	Computer	Mathematics	System-paced	142	Low
	Transfer	−0.45	Personalization	College	English	Computer	Mathematics	System-paced	142	Low
Doolittle (2010)	Retention	0.01	Personalization	College	English	Computer	Social Science	System-paced	3	Mixed
	Transfer	0.16	Personalization	College	English	Computer	Social Science	System-paced	3	Mixed
	Retention	0.13	Personalization	College	English	Computer	Social Science	System-paced	150	Mixed
	Transfer	−0.05	Personalization	College	English	Computer	Social Science	System-paced	150	Mixed
Ginns and Fraser (2010)	Interest	0.42	Personalization	Mixed levels of adult education	English	Paper	Science	System-paced	25	Mixed
	Retention	0.71	Personalization	Mixed levels of adult education	English	Paper	Science	System-paced	25	Mixed
Kartal (2010)	Friendliness	0.41	Personalization	College	Turkish	Computer	Science	Self-paced	Not stated	Mixed
	Learning assistance	0.65	Personalization	College	Turkish	Computer	Science	Self-paced	Not stated	Mixed
	Interest	0.36	Personalization	College	Turkish	Computer	Science	Self-paced	Not stated	Mixed

Table 1 (continued)

Study	Variate type	Standardised mean difference (*d*)	Focus	Educational level	Language	Experimental material format	Broad field of study	Pacing of instruction	Length of learning phase (min)	Prior knowledge level
	Effective processing	1.19	Personalization	College	Turkish	Computer	Science	Self-paced	Not stated	Mixed
	Retention	1.14	Personalization	College	Turkish	Computer	Science	Self-paced	Not stated	Mixed
	Transfer	0.69	Personalization	College	Turkish	Computer	Science	Self-paced	Not stated	Mixed
Stiller and Jedlicka (2010)	Retention	0.87	Personalization	Senior high school	German	Computer	Science	System-paced	13	Low
	Transfer	0.52	Personalization	Senior high school	German	Computer	Science	System-paced	13	Low
	Retention	0.11	Personalization	Senior high school	German	Computer	Science	System-paced	13	High
	Transfer	−0.73	Personalization	Senior high school	German	Computer	Science	System-paced	13	High
McLaren et al. (2011a)	Transfer	0.20	Politeness	Senior high school	English	Computer	Science	Self-paced	90	Low
	Transfer	0.11	Politeness	Senior high school	English	Computer	Science	Self-paced	90	High
McLaren et al. (2011b)	Transfer	0.53	Politeness	Senior high school	English	Computer	Science	Self-paced	90	Low
	Transfer	−0.37	Politeness	Senior high school	English	Computer	Science	Self-paced	90	High
Schworm and Stiller (2012)	Retention	0.24	Personalization	College	German	Computer	Science	System-paced	12	Low
	Transfer	0.83	Personalization	College	German	Computer	Science	System-paced	12	Low

reported results for both the "friendliness" and "helpfulness" of the computer-based tutorial, and results across these measures were averaged into a "perceived friendliness" effect. Lastly, investigations involving multiple experimental conditions compared to a control group required consideration. Kartal (2010) reported results across neutral-formal, personalized-formal and personalized-informal forms of address in Turkish; the first and last conditions were compared in the present analysis on the basis that the personalized-formal condition should be the strongest operationalisation of a conversational style in Turkish. Likewise, Schworm and Stiller (2012) compared formal, "weak" personalization and "strong" personalization of instructional materials written in German, and results from the first and last conditions were included in the present analysis.

Coding of Study Features

The available studies of conversational style vary across a range of factors which might moderate the effects across learning processes and outcomes. To investigate potential moderators, the following study features were coded by the first two authors using a process of consensus coding:

1. Publication type—journal article, conference paper or unpublished thesis/dissertation.
2. Operationalisation of conversational style—personalization or politeness. As only one "visible author" effect was available, this was coded as "personalization".
3. Educational level—junior high school, senior high school, college or mixed levels of adult education.
4. Language of instruction—English vs. other.
5. Form of experimental materials—computer-based vs. paper-based.
6. Broad field of study—mathematics/logic, science, social science, language arts or industrial arts.
7. Pacing of instruction—self-paced vs. system-paced.
8. Length of learning phase.
9. Prior knowledge level of participants—low, medium, high or mixed.

Estimation and Interpretation of Effect Sizes

Following typical convention, the sign of the standardised mean difference (d) effect sizes, calculated using Hedges and Olkin's (1985) small-sample adjustment, given in Table 1 is positive when the use of conversational style had a positive effect (e.g. higher levels of interest, lower ratings of difficulty, higher test scores). Each effect was weighted by its inverse variance weight in order that studies with larger samples contributed more to estimates of population parameters (Hedges and Olkin 1985). For both average effect and moderator analyses, mean sample size-weighted effects and their 95% confidence intervals were estimated using random-effects tests based on the method of moments, instantiated through SPSS syntax provided by Wilson (2010). Whilst meta-analyses based on fixed-effects methods have historically been more prevalent, meta-analyses based on random-effects models support inferences beyond the studies included in the analysis (Hedges and Vevea 1998; Hunter and Schmidt 2000).

For each learning process and outcome, homogeneity in effect sizes was assessed using the Q test as a test of significance, indicating whether the observed variability in effect sizes was unlikely to have arisen by chance. Under conditions of homogeneity of variance, the Q test is distributed as an approximate χ^2 statistic with $k-1$ degree of freedom; when Q

exceeds the critical value of χ^2, the effect sizes under analysis are assumed to vary due to one or more moderating variables.

Publication bias—the tendency for the availability of research to depend on results (Vevea and Woods 2005)—is a common concern when conducting meta-analysis, and a number of methods are available to assess the degree to which this issue might affect conclusions. We evaluate the possibility of publication bias using two separate methods. The first involves inspection of funnel plots where the effect size is plotted against the standard error for each study. Lack of bias is suggested where the data are funnel-shaped and symmetrical. However, the subjectivity inherent in inspections of funnel plots supports the use of additional tests of bias. The second method is a form of sensitivity analysis developed by Vevea and Woods (2005) and instantiated in R syntax described in Field and Gillett (2010). Four weight functions, correcting for moderate one-tailed selection, severe one-tailed selection, moderate two-tailed selection and severe two-tailed selection, are specified in order to examine the extent to which the average effect size changes under each of the above situations.

Our interpretation of the magnitude of effect sizes and associated confidence intervals draws on a major review of educational meta-analyses by Hattie (2009). Based on results across over 800 meta-analyses, Hattie suggested values of *d* of 0.20, 0.40 and 0.60 be used to characterise small, medium and large effects on educational achievement, respectively. We adopt the above benchmarks in interpreting both learning processes (based on self-reports) and outcomes, but note that these are only rough rules of thumb that might not be appropriate in relation to a particular research area.

Results

Tables 2 and 3 present the number of effect sizes, weighted mean effect sizes and confidence intervals within moderator categories for each of the learning processes and outcomes, respectively. Figure 2 shows the weighted mean effect size and confidence interval for each of the learning processes and outcomes.

Hypothesis 1—Perceived Friendliness

The average effect of conversational style on participants' perceptions of friendliness of instructional materials was positive and moderate in size (d=0.46), supporting hypothesis 1; conversational instructional text generates a more social response to learning materials, as indexed by perceptions of friendliness. Plausible values for the effect ranged from small to large positive effects (95% CI=0.22–0.70). The homogeneity statistic was not statistically significant (Q=2.82, df=3, p=0.420), but this result should be viewed with caution given the small number of effects available for analysis. Likewise, tests of publication bias should be viewed with caution; the forest plot was difficult to interpret with only four effects. Tests of bias using Vevea and Woods' (2005) methods found only small departures from the estimated population effect (d=0.46) under moderate one-tailed selection (d=0.43), severe one-tailed selection (d=0.40), moderate two-tailed selection (d=0.43) and severe two-tailed selection (d=0.40).

Table 2 Learning processes: overall results and follow-up moderator analyses (where applicable)

Variable and moderators	No. of studies	Standardised mean difference (*d*)	95% Confidence interval for *d*	
Perceived friendliness (Q=2.82, df=3, p=0.420)	5	0.46	0.22	0.70
Perceived learning assistance (Q=12.36, df=4, p=0.015)	5	0.16	−0.19	0.52
Language of instruction				
English	3	0.03	−0.49	0.55
Other	2	0.35	−0.30	0.99
Publication type				
Journal article	3	0.17	−0.38	0.72
Other	2	0.12	−0.48	0.75
Interest (Q=7.52, df=5, p=0.185)	6	0.15	−0.13	0.44
Effective cognitive processing (Q=9.90, df=3, p=0.019)	4	0.62	0.13	1.11
Language of instruction				
English	2	0.63	−0.22	1.50
Other	2	0.61	−0.21	1.44

Hypothesis 2—Perceived Learning Assistance

The average effect of conversational style on participants' perceptions of learning assistance of instructional materials was small in size (d=0.16). Plausible values for the effect ranged from a small negative effect to a large positive effect (95% CI= −0.19 to 0.52), meaning hypothesis 2—conversational instructional text generates a more social response to learning materials, as indexed by perceptions of learning assistance—was not supported. The homogeneity statistic was statistically significant (Q=12.36, df=4, p=0.015), indicating that one or more moderators conditioned the overall effect. The small number of effects for this variate meant that follow-up moderation analysis was limited due to insufficient variation across study features (e.g. all available effects having the same coding on a given moderator, or where variance did occur across categories, one category having only one effect). A test for moderation by language of instruction (English vs. other) was not statistically significant (Q=0.56, df=3, p=0.455), nor was a test for publication type (journal vs. other; Q=0.08, df=3, p=0.930).

Inspection of the forest plot suggested some asymmetry of results favouring more positive effects. Tests of bias using Vevea and Woods' (2005) methods found only small departures from the estimated population effect (d=0.16) under moderate one-tailed selection (d=0.09), moderate two-tailed selection (d=0.16) and severe two-tailed selection (d= 0.15). However, there was a major discrepancy under severe one-tailed selection (d=−1.05).

Table 3 Learning outcomes: overall results and follow-up moderator analyses (where applicable)

Variable and moderators	No. of studies	Standardised mean difference (*d*)	95% Confidence interval for *d*	
Retention ($Q = 54.05$, $df = 29$, $p = 0.003$)	30	0.30	0.18	0.41
Language of instruction				
English	25	0.25	0.13	0.36
Other	5	0.55	0.27	0.82
Publication type				
Journal article	20	0.38	0.24	0.53
Conference paper	2	0.03	−0.37	0.42
Thesis/dissertation	8	0.21	0.03	0.39
Operationalisation				
Personalization	28	0.29	0.18	0.41
Politeness	2	0.38	−0.13	0.89
Educational level				
Senior high school	2	0.49	0.04	0.93
College	26	0.28	0.16	0.40
Mixed adult education	2	0.39	−0.11	0.90
Form of materials				
Computer-based	26	0.26	0.15	0.38
Paper-based	4	0.50	0.20	0.80
Broad field of study				
Maths/logic, science and engineering	25	0.30	0.17	0.43
Social science and language arts	5	0.28	0.02	0.53
Pacing of instruction				
Self-paced	15	0.22	0.05	0.39
System-paced	14	0.35	0.19	0.50
Prior knowledge level				
Low	8	0.25	0.01	0.49
Mixed	16	0.39	0.21	0.56
High	3	0.30	−0.11	0.71
Instruction time				
<5 min	6	0.07	−0.18	0.32
10–20 min	6	0.37	0.17	0.58
20–35 min	6	0.55	0.29	0.81
≥35 min	4	−0.01	−0.28	0.25
Transfer ($Q = 172.18$, $df = 24$, $p < 0.001$)	25	0.54	0.25	0.83
Language of instruction				
English	19	0.62	0.33	0.90
Other	6	0.12	−0.39	0.63
Publication type				
Journal article	22	0.55	0.27	0.82
Conference paper or thesis/dissertation	3	0.20	−0.51	0.91
Operationalisation				
Personalization	21	0.58	0.31	0.86

Table 3 (continued)

Variable and moderators	No. of studies	Standardised mean difference (*d*)	95% Confidence interval for *d*	
Politeness	4	0.11	−0.49	0.72
Educational level				
Junior high school	2	−0.03	−0.94	0.87
Senior high school	4	0.04	−0.58	0.65
College	19	0.65	0.37	0.94
Broad field of study				
Maths/logic, science and industrial design	22	0.55	0.28	0.82
Social science	3	0.18	−0.52	0.88
Pacing of instruction				
Self-paced	11	0.43	0.06	0.81
System-paced	13	0.63	0.29	0.98
Prior knowledge level				
Low	10	0.60	0.20	1.00
Medium	2	−0.11	0.97	0.74
Mixed	11	0.66	0.28	1.04
High	2	−0.28	−1.16	0.71
Instruction time				
<5 min	6	0.79	0.29	1.30
10–20 min	6	0.45	−0.05	0.95
20–35 min	2	1.41	0.52	2.31
≥35 min	4	−0.05	−0.63	0.53

Hypothesis 3—Interest

The average effect of conversational style on participants' interest in the topic was small in size (d=0.15), and plausible magnitudes for the effect ranged from a small negative effect to a large positive effect (95% CI=−0.13 to 0.44); thus, hypothesis 3—conversational instructional text generates higher levels of interest—was not supported. The homogeneity statistic was not statistically significant (Q=7.52, df=5, p=0.185), but this result should be viewed with caution given the small number of effects available for analysis.

Inspection of the forest plot suggested some asymmetry of results favouring more positive effects. Tests of bias using Vevea and Woods' (2005) methods found only small departures from the estimated population effect (d=0.15) under moderate one-tailed selection (d=0.07), moderate two-tailed selection (d=0.12) and severe two-tailed selection (d=0.10). However, there was a major discrepancy under severe one-tailed selection (d=−0.11).

Hypothesis 4—Effective Cognitive Processing

The average effect of conversational style on participants' perceptions of effective cognitive processing was positive and large in size (d=0.62), supporting hypothesis 4; conversational

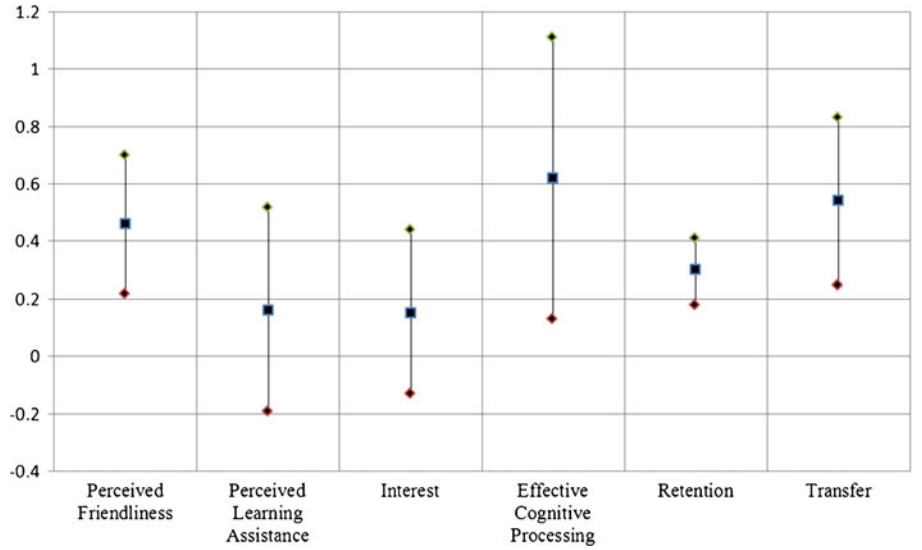

Fig. 2 Weighted mean effect size and confidence interval for learning process and outcome variates

text supports more effective cognitive processing, as indexed by lower ratings of difficulty during the learning phase. Plausible values for the effect ranged from very small to large positive effects (95% CI=0.13–1.11). The homogeneity statistic was statistically significant (Q=9.90, df=3, p=0.019), indicating that one or more moderators conditioned the overall effect. As for the learning assistance moderation tests, the small number of effects for this variate limited follow-up moderation analyses. A test for moderation by language of instruction (English vs. other) was not statistically significant (Q=0.001, df=3, p=0.973).

Inspection of the forest plot suggested some asymmetry of results favouring more positive effects. Tests of bias using Vevea and Woods' (2005) methods found only small departures from the estimated population effect (d=0.62) under moderate one-tailed selection (d=0.54), moderate two-tailed selection (d=0.57) and severe two-tailed selection (d=0.52), but a somewhat stronger discrepancy under severe one-tailed selection (d=0.46).

Hypothesis 5—Tests of Retention

The average effect of conversational style on participants' performance on tests of retention was positive and between small and medium in size (d=0.30), supporting hypothesis 5; conversational text supports more effective learning, as indexed by students' recall of content covered during instruction. Plausible values for the effect ranged from small to medium positive effects (95% CI=0.18–0.41). The homogeneity statistic was statistically significant (Q=54.05, df=29, p=0.003), indicating that one or more moderators conditioned the overall effect. The test for moderation by language of instruction was statistically significant (Q=3.94, df=28, p=0.047): on average, studies carried out in English had lower effect sizes (d= 0.25, 95% CI=0.13–0.36) than those carried out in other languages (d=0.55, 95% CI=0.27–0.82). A test for moderation by publication type (journal article, conference paper or thesis/dissertation) was not statistically significant (Q=3.95, df=27, p=0.139), nor was a test for operationalisation (personalization vs. politeness; Q=0.10, df=28, p=0.752). The test for moderation by educational level (senior high school, college or mixed levels of

adult education) was also not statistically significant (Q=0.93, df=27, p=0.628). The form of instructional materials (computer-based vs. paper-based) did not moderate the overall effect (Q=2.02, df=28, p=0.155).

The broad field of study focus of the available effects for retention was skewed towards science (23/30 studies), and a number of categories (maths/logic, engineering, language arts) only had one effect. Thus, categories for broad field of study were collapsed into two categories: maths/logic, science and engineering (k=25); and social science and language arts (k=5). This broad categorisation did not moderate the overall effect (Q=0.03, df=28, p=0.858), nor did the pacing of instruction (self-paced vs. system-paced; Q=1.07, df=27, p=0.302). For prior knowledge level of participants, only one effect (Yeung et al. 2009; study 3) was coded as "medium", and so this effect was excluded from the analysis. The resulting test of moderation (low, mixed and high levels of prior knowledge) was not statistically significant (Q=0.84, df=24, p=0.657).

Lastly, to examine whether the average effect was moderated by the length of instruction time, the distribution of instruction periods was examined. As a continuous rather than categorical variable, under ideal circumstances, time on instruction might be examined using meta-analytic regression; however, the relatively small number of effects, combined with a substantial number of missing data points for these variates, militated against this strategy. For both retention and transfer effects, break points in the distributions of available effects suggested four suitable categories: <5 min, 10–20 min, 20–35 min and 35 min or greater. The resulting test for moderation was statistically significant (Q=12.49, df=18, p=0.006): on average, studies based on 5 min or less instruction time had very small effect sizes (d=0.07, 95% CI=−0.18 to 0.32) compared to those based on 10–20 min of instruction time (d=0.37, 95% CI=0.17–0.58) or 20–35 min (d=0.55, 95% CI=0.29–0.81). Studies based on more than 35 min of instruction time also had very low effect sizes (d=−0.02, 95% CI=−0.28 to 0.25). These results should, however, be viewed with caution given the substantial proportion (28 %) of effects with missing values on this variable.

Inspection of the forest plot suggested some asymmetry of results favouring more positive effects. Tests of bias using Vevea and Woods' (2005) methods found only small departures from the estimated population effect (d=0.30) under moderate one-tailed selection (d=0.23), moderate two-tailed selection (d=0.26) and severe two-tailed selection (d=0.23), but a more substantial discrepancy under severe one-tailed selection (d=0.13).

Hypothesis 6—Tests of Transfer

The average effect of conversational style on participants' performance on tests of transfer was positive and moderate to large in size (d=0.54), supporting hypothesis 6; conversational text supports more effective learning, as indexed by students' capacity to transfer learning to novel circumstances. Plausible values for the effect ranged from small to large positive effects (95% CI=0.25–0.83). The homogeneity statistic was statistically significant (Q=172.18, df=24, p<0.001), indicating that one or more moderators conditioned the overall effect.

A test for moderation by language of instruction was not statistically significant (Q=2.77, df=23, p=0.096). The test of moderation by publication type was based on a comparison of journal article publication vs. combined other categories (conference paper or thesis/dissertation), but was not statistically significant (Q=0.81, df=23, p=0.368), nor was a test for operationalisation (personalization vs. politeness; Q=1.92, df=23, p=0.166). The test for moderation by educational level (junior high school, senior high school or college students) was also not statistically significant (Q=4.64, df=22, p=0.098). A test of

moderation by form of instructional materials (computer-based vs. paper-based) could not be conducted as all but one effect were coded as computer-based.

The broad field of study focus of the available effects for retention was skewed towards science (20/25 studies), and two categories (maths/logic and industrial design) only had one effect. Thus, categories for broad field of study were collapsed into two categories: maths/logic, science and industrial design (k=22); and social science (k=3). This broad categorisation did not moderate the overall effect (Q=0.93, df=23, p=0.335), nor did the pacing of instruction (self-paced vs. system-paced; Q=0.59, df=22, p=0.443). The test of moderation for prior knowledge level (low, medium, mixed and high levels of prior knowledge) was not statistically significant (Q=5.23, df=21, p=0.156).

Using the four categories for length of instruction time described above, the resulting test for moderation was statistically significant (Q=8.68, df=14, p=0.034). The pattern across categories was unclear: on average, studies based on 5 min or less of instruction time had a large average effect size (d=0.79, 95% CI=0.29–1.30) compared to a smaller and statistically unreliable effect for those based on 10–20 min of instruction time (d=0.45, 95% CI=−0.05 to 0.95). Studies based on 20–35 min had a very large average effect (d=1.41, 95% CI=0.52–2.31), whilst those based on more than 35 min of instruction time on average had a small, slightly negative but statistically unreliable effect sizes (d=−0.05, 95% CI=−0.63 to 0.53). As for the retention results, these results should be viewed with caution given the substantial proportion (25 %) of effects with missing values.

Inspection of the forest plot suggested substantial asymmetry of results favouring more positive effects. Tests of bias using Vevea and Woods' (2005) methods found a moderate departure from the estimated population effect (d=0.54) under moderate one-tailed selection (d=0.35) and severe two-tailed selection (d=0.42), but less of a departure under moderate two-tailed selection (d=0.49). The departure under severe one-tailed selection (d=−1.56) was very large.

Discussion and Conclusions

Summary and Discussion

Drawing on Mayer's (2009) theorising, the present meta-analysis tested hypotheses that the use of conversational rather than formal style in instructional text would generate positive effects across a range of learning outcomes (retention and transfer) and learning processes (perceived friendliness, perceived learning assistance, interest, effective cognitive processing). Consistent with Mayer's model, students who learned from instructional materials written in a conversational style learned more from instruction than those who studied more formally expressed materials. Notably, reliable effects on learning were found not only on tests of retention (a small to medium overall effect, d=0.30) but also on tests of transfer (d=0.54, a moderate to large overall effect), consistent with Mayer's (2009) argument that a conversational style promotes deeper learning than a formal style. Across studies, participants rated the materials as being more friendly (d=0.46) as well as promoting more effective cognitive processing during a learning phase (d=0.62). However, effects of a conversational style on perceived learning assistance and interest were smaller, on average (d=0.16 and 0.15, respectively), and were not statistically reliable.

Whilst many learning events may appear on face value to be solitary, Mayer (2009) has suggested that such events may in fact be inherently social, to the extent that an "implied conversation" may be taking place between the instructor and the learner. This possibility

suggests that some instructional designs may be more effective than others in implying such a conversation. A number of perspectives, including schema theory (Spiro 1977), narrative text research (Graesser et al. 2002) and research on the dynamics of human–computer interactions (Reeves and Nass 1996), provided the foundation for the body of instructional design research reviewed here.

The above analyses included tests of generalizability across a range of potential moderators, such as field of study, educational level and presentation format. Tests of heterogeneity were significant for two learning processes—perceived learning assistance and effective processing—and for the two learning outcomes under analysis: retention and transfer. Follow-up tests of moderation were only partially successful in identifying factors conditioning these effects, often because the effects came from subsets of studies which were themselves quite homogenous in their features. Thus, no factor could be identified from those coded that moderated learning assistance or effective cognitive processing effects. Overall, the small number of learning process effect sizes available for analysis constitutes an inherent limitation to understanding potential boundary conditions across these variables.

The larger number of effects for the learning outcomes of retention and transfer yielded more scope for identifying moderators, but across these variates, only two significant moderators were identified. For retention, studies conducted in English evinced lower effect sizes than those conducted in other languages (German, Flemish or Turkish); however, language did not moderate effect sizes for transfer. Further studies across a range of languages would support a clearer understanding of the effects of conversational style. In addition, the existence within some languages of multiple modes of address varying across a continuum of formality (e.g. Turkish; see Kartal 2010) provides the opportunity for quite "fine-grained" investigations of conversational effects.

Across all variates, the clearest indication of moderation was by length of instruction time. Earlier investigations of the personalization principle by Mayer and colleagues (e.g. Mayer et al. 2004; Moreno and Mayer 2000, 2004) used computer-based tutorials of relatively short duration, typically <5 min. Subsequent studies (e.g. Doolittle 2010; Ginns and Fraser 2010) explored whether the effects found in earlier studies would carry over to longer study periods. Across the experiments for which length of instruction time was stated, the benefits of conversational style varied substantially. Conversational style was effective in studies of <5-min duration as measured on tests of transfer, but not retention. For studies lasting between 10 and 20 min, conversational-style redesigns were effective as measured on tests of retention, but not transfer. However, considering the mostly positive values of the confidence interval for the transfer mean effect, future studies might focus on this duration of instruction in order to clarify this estimate of the population effect. For studies lasting between 20 and 35 min, conversational style was effective both on tests of retention and transfer. Lastly, for studies lasting between more than 35 min, conversational style did not improve retention or transfer.

Taken together, these results suggest a boundary condition for redesigns using conversational style: positive effects of conversational style on learning tend to be seen more clearly in materials studied for a half hour or less, after which the effects are close to zero. The reasons for this apparent boundary condition are at present unclear; possibly, across longer instructional periods, the relative novelty of instructional materials written in conversational style becomes normalised, leading to the effects on cognitive processing becoming less potent.

Alternatively, these results may be explicable if students' interactions with paper- or computer-presented materials are cast as a quasi-social interaction with a "virtual teacher". Reviewing social information processing theory (Walther 1992) and related research on text-

based computer-mediated communication (CMC), Antheunis et al. (2012) note that conversation partners can overcome the lack of non-verbal cues in CMC in *initial* text-based conversations through increased use of uncertainty reduction strategies such as question asking and self-disclosure (e.g. Ramirez et al. 2002; Tidwell and Walther 2002), strategies used in the conversational-style redesigns reviewed above. In the context of an "instructional conversation" with a text-based virtual teacher, then, uncertainty reduction strategies may decline in potency the longer the conversation progresses; there may only be so much uncertainty that can be reduced before further attempts appear redundant.

Limitations and Future Directions

The findings of this meta-analysis provide some support for the theoretical model of conversational-style processes and effects given in Fig. 1, insofar as hypotheses for perceived friendliness, effective cognitive processing and learning outcomes were supported. However, hypotheses regarding perceived learning assistance and interest were not supported. Given the relatively small number of studies which examined these hypothesised learning processes, and the results in line with hypotheses for other learning processes, it is possible that the current results represent type 2 errors. Further studies are needed where a range of both learning process and outcome measures, including the "gold standard" of learning transfer, are used—supporting direct comparisons of the different effects—in order to build on the results of this review. Investigators may also need to consider the design of such measures carefully. Many conversational style investigations employing learning process measures have used single-item ratings administered at the end of instruction; such measures may not be sufficiently sensitive to detect real group differences, particularly when the period of instruction is longer. Future studies might obtain more sensitive estimates by asking participants to report on hypothesised processes at multiple points during learning rather than make overall ratings after the learning phase (for examples using interest ratings, see Ainley et al. 2005, 2002).

Other methodologies might also be considered to investigate cognitive processing hypotheses related to conversational style. Johnson and Mayer (2012) used eye movement analysis to test cognitive processing during learning from spatially contiguous vs. non-contiguous words and diagrams. Across three experiments, they found large effects favouring the contiguous condition in the number of integrative transitions (eye movements across text and diagrams) and corresponding transitions (eye movements from text to the corresponding section of the diagram), as well as on two of three tests of transfer. The use of conversational style might similarly promote more effective processing across text and diagrams as a student's attention is focused periodically by reference to him/herself and/or the use of narrative features. The degree to which visual processing is affected when conversational-style materials are presented in an auditory format would also be worthy of investigation.

The studies reviewed in this meta-analysis varied across a number of dimensions. The diversity of studies—ranging across operationalisation of conversational style, educational topic, educational level, presentation media and presentation length—is a sign of the "health" of this evidence base as investigators seek to replicate and extend prior research. The results of this review suggest a number of considerations for researchers working in this area. Firstly, we advocate the inclusion of learning process measures in future studies as well as learning outcomes, with particular consideration of their measurement properties. Secondly, we recommend that future studies investigate conversational-style effects across longer (e.g. more than 30 min) periods of

instruction to determine more clearly whether the apparent boundary condition identified in the present meta-analysis is reliable. Following the discussion above of the role of uncertainty reduction in computer-mediated conversations, more fine-grained repeated measures of learning could test the hypothesis that conversational-style redesigns will improve learning of material covered in the earlier stages of instruction to a greater extent than materials covered in later stages. Such studies might also investigate whether "interleaving" of conversational and non-conversational materials of different lengths (e.g. at the 30-min mark) might promote attentional capture and subsequent learning. Thirdly, investigators might take the categories of studies listed in Tables 2 and 3 as starting points for identifying gaps in the literature. For instance, the majority of studies were conducted in English, but there is some evidence that linguistic forms in other languages may vary across a continuum of conversational style (e.g. Kartal 2010); more investigations across other languages would be valuable in establishing the cross-cultural generalizability of this effect. Because of the small number of studies conducted in languages other than English, the present meta-analysis was limited in its ability to investigate this moderator; nonetheless, language of instruction moderated effects on retention, suggesting that the efficacy of conversational style may be stronger in some languages than others. Moreover, our review identified no studies of the politeness effect that have been conducted in languages other than English. A substantial evidence base exists on the variation in politeness norms across cultures (Bargiela-Chiappini and Kadar 2010) that could form the basis for future hypothesis generation and testing.

Fourthly, larger sample sizes will support more powerful tests of moderation of main effects by variables such as prior knowledge. The capacity of this meta-analysis to investigate moderation by knowledge level was often limited by vague reporting of prior knowledge levels of participants as well as the small number of studies where effects of different knowledge level were tested (e.g. McLaren et al. 2011b). Larger sample sizes, taking in a range of reliably measured knowledge levels, will support more sensitive tests of moderation (e.g. the Johnson–Neyman method; for a review, see Hayes and Matthes 2009) than the median-split method used in some studies (for critiques of this method, see Cohen 1983; MacCallum et al. 2002).

Conclusion

This meta-analysis reviews the impact of instructional redesigns aimed at making text more conversational. Underpinning the rationale for each of these studies is the expectation that instructional text written in a more conversational style of writing will lead to better learning outcomes than text written in a more formal style. These redesigns have been operationalized in a number of ways, including making changes from third person to first and second person; adding sentences which directly address the reader; using forms of address that are more polite; and making the author's views and personality more visible. Across studies identified for meta-analysis, the average impact of such redesigns on measures of retention and transfer was moderate to large across a range of topics, educational levels and instructional media. These results should be of substantial interest to designers of instructional materials as they result from relatively modest changes to instructional materials. However, the meta-analysis also revealed substantial scope for additional research in this area. In particular, open questions remain about the causal processes underlying conversational-style effects as well as possible boundary conditions.

References

Ainley, M., Corrigan, M., & Richardson, N. (2005). Students, tasks and emotions: identifying the contribution of emotions to students' reading of popular culture and popular science texts. *Learning and Instruction, 15*, 433–447. doi:10.1016/j.learninstruc.2005.07.011.

Ainley, M., Hidi, S., & Berndorff, D. (2002). Interest, learning and the psychological processes that mediate their relationship. *Journal of Educational Psychology, 94*, 545–561. doi:10.1037//0022-0663.94.3.545.

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., et al. (2001). *A taxonomy for learning, teaching, and assessing: a revision of Bloom's taxonomy of educational objectives*. New York: Longman.

Antheunis, M. L., Schouten, A. P., Valkenburg, P. M., & Peter, J. (2012). Interactive uncertainty reduction strategies and verbal affection in computer-mediated communication. *Communication Research, 39*, 757–780. doi:10.1177/0093650211410420.

Bargiela-Chiappini, F., & Kadar, D. (Eds.). (2010). *Politeness across cultures*. Basingstoke: Palgrave.

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Hillsdale, NJ: Lawrence Erlbaum.

Bretzing, B. H., & Kulhavy, R. W. (1981). Note-taking and passage style. *Journal of Educational Psychology, 73*, 242–250.

Brown, P., & Levinson, S. C. (1987). *Politeness: some universals in language use*. New York, NY: Cambridge University Press.

Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement, 7*, 249–253.

Clarebout, G., & Elen, J. (2007). In search of pedagogical agents' modality and dialogue effects in open learning environments. *e-Journal of Instructional Science and Technology (e-JIST)*, 10. Retrieved 3 September 2010 from http://www.ascilite.org.au/ajet/e-jist/docs/vol10_no1/papers/full_papers/clarebout_elen.htm.

D'Ailly, H. H., Murray, H. G., & Corkill, A. (1995). Cognitive effects of self-referencing. *Contemporary Educational Psychology, 20*, 88–113.

Doolittle, P. (2010). The effects of segmentation and personalization on superficial and comprehensive strategy instruction in multimedia learning environments. *Journal of Educational Multimedia and Hypermedia, 19*, 159–175.

Dunsworth, Q. (2005). Fostering multimedia learning of science: the role of personalization and presentation mode. Doctoral dissertation. Available from ProQuest Dissertations and Theses database. UMI no. AAT 3173234.

Field, A. P., & Gillett, R. (2010). How to do a meta-analysis. *British Journal of Mathematical and Statistical Psychology, 63*, 665–694. doi:10.1348/000711010X502733.

Ginns, P., & Fraser, J. (2010). Personalization enhances learning anatomy terms. *Medical Teacher, 32*, 776–778. doi:10.3109/01421591003692714.

Graesser, A. C., Hauft-Smith, K., Cohen, A. D., & Pyles, L. D. (1980a). Advanced outlines, familiarity, text genre, and retention of prose. *The Journal of Experimental Education, 48*, 209–220.

Graesser, A. C., Hoffman, N. L., & Clark, L. F. (1980b). Structural components of reading time. *Journal of Verbal Learning and Verbal Behavior, 19*, 131–151.

Graesser, A. C., Olde, B., & Klettke, B. (2002). How does the mind construct and represent stories? In M. C. Green, J. J. Strange, & T. C. Brock (Eds.), *Narrative impact: social and cognitive foundations* (pp. 231–263). Mahwah, NJ: Lawrence Erlbaum.

Graesser, A. C., & Ottati, V. (1996). Why stories? Some evidence, questions, and challenges. In R. S. Wyer (Ed.), *Knowledge and memory: the real story* (pp. 121–132). Hillsdale, NJ: Lawrence Erlbaum.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics 3: Speech arts* (pp. 41–58). New York, NY: Academic.

Haberlandt, K., & Graesser, A. C. (1985). Component processes in text comprehension and some of their interactions. *Journal of Experimental Psychology. General, 114*, 357–374.

Hattie, J. (2009). *Visible learning: a synthesis of over 800 meta-analyses relating to achievement*. New York, NY: Routledge.

Hayes, A. F., & Matthes, J. (2009). Computational procedures for probing interactions in OLS and logistic regression: SPSS and SAS implementations. *Behavior Research Methods, 41*, 924–936. doi:10.3758/BRM.41.3.924.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. New York: Academic Press.

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods, 3*, 486–504. doi:10.1037/1082-989X.3.4.486.

Hunter, J. E., & Schmidt, F. L. (2000). Fixed effects vs. random effects meta-analysis methods: Implications for cumulative research knowledge. *International Journal of Selection and Assessment, 8*, 275-292. doi:10.1111/1468-2389.00156.

Inglese, T., Mayer, R. E., & Rigotti, F. (2007). Using audiovisual TV interviews to create visible authors that reduce the learning gap between native and non-native speakers. *Learning and Instruction, 17*, 67–77. doi:10.1016/j.learninstruc.2006.11.006.

Johnson, C. I., & Mayer, R. E. (2012). An eye movement analysis of the spatial contiguity effect in multimedia learning. *Journal of Experimental Psychology: Applied, 18*, 178–191. doi:10.1037/a0026923.

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review, 19*, 509–539. doi:10.1007/s10648-007-9054-3.

Kartal, G. (2007). How universal are e-learning design guidelines? Reconsidering the personalization principle. In D. Remenyi (Ed.), *Proceedings of the 2nd International Conference on E-Learning* (pp. 269–275). Reading, UK: Academic Conferences.

Kartal, G.. (2010). Does language matter in multimedia learning? Personalization principle revisited. *Journal of Educational Psychology*, *102*, 615–624. doi:10.1037/a0019345.

Lorch, R. F. (1989). Text-signaling devices and their effects on reading and memory processes. *Educational Psychology Review, 1*, 209–234. doi:10.1007/BF01320135.

MacCallum, R.C., Zhang, S.,. Preacher, K.J., & Rucker, D.D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7*, 19-40. doi:10.1037//1082-989X.7.1.19.

Mayer, R. E. (2005a). Principles of multimedia learning based on social cues: personalization, voice, and image principles. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 201–212). Cambridge: Cambridge University Press.

Mayer, R. E. (2008). Applying the science of learning: evidence-based principles for the design of multimedia instruction. *American Psychologist, 63*, 760–769. doi:10.1037/0003-066X.63.8.770b.

Mayer, R. E. (2001). *Multimedia learning* (1st ed.). New York, NY: Cambridge University Press.

Mayer, R. E. (2005b). Principles of multimedia learning based on social cues: personalization, voice, and image principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 201–212). New York, NY: Cambridge University Press.

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). New York, NY: Cambridge University Press.

Mayer, R. E., Fennell, S., Farmer, L., & Campbell, J. (2004). A personalization effect in multimedia learning: students learn better when words are in conversational style rather than formal style. *Journal of Educational Psychology, 96*, 389–395. doi:10.1037//1082-989X.7.1.19.

Mayer, R. E., Johnson, W. L., Shaw, E., & Sandhu, S. (2006). Constructing computer-based tutors that are socially sensitive: politeness in educational software. *International Journal of Human Computer Studies, 64*, 36–42. doi:10.1016/j.ijhcs.2005.07.001.

McLaren, B. M., DeLeeuw, K. E., & Mayer, R. E. (2011a). Polite web-based intelligent tutors: can they improve learning in classrooms? *Computers & Education, 56*, 574–584. doi:10.1016/j.compedu.2010.09.019.

McLaren, B. M., DeLeeuw, K. E., & Mayer, R. E. (2011b). A politeness effect in learning with web-based intelligent tutors. *International Journal of Human-Computer Studies, 69*, 70–79.

McLaren, B. M., Lim, S., Gagnon, F., Yaron, D., & Koedinger, K. R. (2006). Studying the effects of personalized language and worked examples in the context of a web-based intelligent tutor. *Proceedings of the 8th International Conference on Intelligent Tutoring Systems*, Jhongli, Taiwan, June 26–30, 2006.

McLaren, B.M., Lim, S., Yaron, D., & Koedinger, K.R. (2007). Can a polite intelligent tutoring system lead to improved learning outside of the lab? In R. Luckin, K.R. Koedinger, & J. Greer (Eds.), *Proceedings of the 13th International Conference on Artificial Intelligence in Education* (AIED-07), Artificial Intelligence in Education: Building Technology Rich Learning Contexts that Work (pp. 433–440). Amsterdam: IOS Press.

Menke, D. J., & Pressley, M. (1994). Elaborative interrogation: using "why" questions to enhance the learning from text. *Journal of Reading, 37*, 642–645. http://www.jstor.org/stable/20172385.

Moreno, R., & Mayer, R. E. (2000). Engaging students in active learning: the case for personalized multimedia messages. *Journal of Educational Psychology, 92*, 724–733. doi:10.1037/0022-0663.92.4.724.

Moreno, R., & Mayer, R. E. (2004). Personalized messages that promote science learning in virtual environments. *Journal of Educational Psychology, 96*, 165–173. doi:10.1037/0022-0663.96.1.165.

Nolen, S. B. (1995). Effects of a visible author in statistics texts. *Journal of Educational Psychology, 87*, 47–65.

Paxton, R. J. (1997). "Someone with like a life wrote it": the effects of a visible author on high school history students. *Journal of Educational Psychology, 89*, 235–250.

Paxton, R. J. (1999). A deafening silence: history textbooks and the students who read them. *Review of Educational Research, 69*, 315–339. doi:10.3102/00346543069003315.

Paxton, R. J. (2002). The influence of author visibility on high school students solving a historical problem. *Cognition and Instruction, 20*, 197–248.

Symons, C., & Johnson, B. (1997). The self-reference effect in memory: a meta-analysis. *Psychological Bulletin, 121*, 371–394.

Ramirez, A., Walther, J. B., Burgoon, J. K., & Sunnafrank, M. (2002). Information-seeking strategies, uncertainty, and computer-mediated communication: toward a conceptual model. *Human Communication Research, 28*, 213–228. doi:10.1111/j.1468-2958.2002.tb00804.x.

Reeves, B., & Nass, C. (1996). *The media equation: how people treat computers, television, and new media like real people and places*. New York, NY: Cambridge University Press.

Robertson, E. (2008). Effects *of personalising anatomy instructions on learning*. Honours thesis, The University of Sydney.

Son, J.Y., & Goldstone, R.L. (2009). Contextualization in perspective. *Cognition and Instruction*, *27*, 51–89. doi:10.1080/07370000802584539.

Spiro, R. J. (1977). Remembering information from text: the "state of schema" approach. In R. C. Anderson, R. J. Spiro, & W. E. Montague (Eds.), *Schooling and the acquisition of knowledge* (pp. 137–165). Hillsdale, NJ: Lawrence Erlbaum.

Stiller, K. D., & Jedlicka, R. (2010). A kind of expertise reversal effect: personalization effect can depend on domain-specific prior knowledge. *Australasian Journal of Educational Technology, 26*, 133–149.

Schworm, S., & Stiller, K. D. (2012). Does personalization matter? The role of social cues in instructional explanations. *Intelligent Decision Technologies, 6*, 105–111. doi:10.3233/IDT-2012-0127.

Tidwell, L. C., & Walther, J. B. (2002). Computer-mediated communication effects on disclosure, impressions, and interpersonal evaluations: getting to know one another a bit at a time. *Human Communication Research, 28*, 17–348.

Vevea, J. L., & Woods, C. M. (2005). Publication bias in research synthesis: sensitivity analysis using a priori weight functions. *Psychological Methods, 10*, 428–443. doi:10.1037/1082-989X.10.4.428.

Wagner, L., Davis, S., & Handelsman, M. M. (1998). In search of the abominable consent form: the impact of readability and personalization. *Journal of Clinical Psychology, 54*, 115–120. doi:10.1002/%28SICI%291097-4679%28199801%2954:1%3C115::AID-JCLP13%3E3.0.CO;2-N.

Walther, J. B. (1992). Interpersonal effects in computer-mediated interaction: a relational perspective. *Communication Research, 19*, 52–90. doi:10.1177/009365092019001003.

Wang, N., & Johnson, W. L. (2008). *The politeness effect in an intelligent foreign language tutoring system. In B. P. Woolf, E. Aimeur, R. Nkambou, and Lajoie, S. (Eds.), Proceedings of the Ninth International Conference on Intelligent Tutoring Systems. Lecture Notes in Computer Science, vol. 5091 (pp. 260–280)*. Berlin: Springer.

Wang, N., Johnson, W.L., Mayer, R.E., Rizzo, P., Shaw, E., & Collins, H. (2008). The politeness effect: pedagogical agents and learning outcomes. *International Journal of Human–Computer Studies*, *66*, 98–112. doi:10.1016/j.ijhcs.2007.09.003

Wilson, D.B. (2010). *SPSS, Stata, and SAS macros for performing meta-analytic analyses.* Retrieved September 19, 2012 from http://mason.gmu.edu/~dwilsonb/ma.html.

Yeung, A., Schmid, S., George, A. V., & King, M. M. (2009). Using the personalization hypothesis to design e-learning environments. In M. Gupta-Bhowon, S. Jhaumeer-Laulloo, H. L. K. Wah, & P. Ramasami (Eds.), *Chemistry education in the ICT Age* (pp. 287–300). Berlin: Springer. doi:10.1007/978-1-4020-9732-4_25.

The following form part of the meta-analysis: Bretzing and Kulhavy (1981); Clarebout and Elen (2007); Doolittle (2010); Dunsworth (2005); Ginns and Fraser (2010); Kartal (2007, 2010); Mayer et al. (2004); McLaren et al. 2006, 2011a, b; Moreno and Mayer 2000, 2004; Paxton (2002); Robertson (2008); Son and Goldstone (2009); Stiller and Jedlicka (2010); Schworm and Stiller (2012); Wagner et al. (1998); Wang and Johnson (2008); Wang et al. (2008); Yeung et al. (2009)