Ambition Institute



Do advance organisers help teachers learn new information during online professional development?

Stefanie Meliss, Christopher Such, Sarah Cottinghatt, Sonal Patel and Sam Sims

2024

Contents

Contributions 3

Acknowledgements 3

Foreword 4

Abstract 5

Do advance organisers help teachers learn new information during online professional development? 6

Methods 9

- Participants 9
- Study design 11
- Materials and measurements 12
- Professional development programme and testing procedures 14
- Data analysis 16

Results 17

- Primary hypothesis testing 17
- Exploratory analysis 19
- Learning experience questionnaire 21

Discussion 22

References 24

Supplementary information 26

- Supplementary methods 26
- Supplementary materials 28
- Supplementary figures 34
- Supplementary tables 39

Contributions

Conceptualisation

Stefanie Meliss
Sarah Cottinghatt
Sam Sims

Data curation

Stefanie Meliss

Formal analysis

Stefanie Meliss

Investigation

Stefanie Meliss
Sonal Patel

Methodology

Stefanie Meliss
Christopher Such
Sarah Cottinghatt
Sam Sims

Project administration

Stefanie Meliss
Sonal Patel
Sam Sims

Software

Stefanie Meliss
Sonal Patel

Supervision

Sam Sims

Validation

Stefanie Meliss

Visualisation

Stefanie Meliss

Writing – original draft

Stefanie Meliss

Writing – review and editing

Christopher Such
Sarah Cottinghatt
Sonal Patel
Sam Sims

Acknowledgements

We thank all participants who donated their valuable time to make this research project possible. We are grateful to Josh Goodrich and the team at Steplab for implementing the technology to enable us to run a randomised experiment on their learning platform. We would further like to acknowledge Tom Parsons and Allie Heny for their contributions in liaising with Steplab to support this development. We thank David Ruddle, Youssoup Pachaev and Unai Vicente for their support in accessing the relevant data from the data warehouse. We further recognise the valuable role Katie McHugh played in advising us on matters of user experience. We express our gratitude to Abigail Brown who supported the management of this project. We are thankful to our colleagues at Ambition Institute, especially Alison Woodrow and Ellie Miles, who supported the communication with the participants. We are grateful to the supervision we received from Jennifer Barker and Marie Hamer. The project was funded by Ambition Institute and funding for this project was acquired by Hilary Spencer and Marie Hamer. Finally, special thanks to Mike Hobbiss for reviewing this paper, and to Hannah Cheetham-Joshi and Briony Banks for their helpful comments on earlier drafts.

Foreword

In the current education landscape, where online professional development has become increasingly prevalent, it is crucial to explore various strategies that can enhance teachers' learning experiences and outcomes. This research paper investigates one such strategy: the use of advance organisers in online professional development settings.

While the use of advance organisers in the classroom has been researched for a long time, there has been little work so far on how they might support teacher professional development. This study presents a nuanced picture of their effectiveness, and seeks to identify some potential areas of future study within the specific context of remote professional development.

Importantly, this research highlights the complexities and challenges inherent in applying academic evidence to practical teaching scenarios. It also underscores the necessity for transparency and specificity in research methodologies to enable practitioners to effectively translate findings into practice.

As we navigate the evolving landscape of teacher education, this study serves as a critical reminder of the importance of ongoing research and adaptation. It encourages educators to remain open to new evidence, to rigorously test new approaches, and to prioritise methods that are demonstrably effective in fostering teacher learning and, ultimately, student success.

I am pleased to present this comprehensive analysis from Ambition Institute. It is the first step in gaining a deeper understanding of both the nature of effective AOs and the ways in which they interact with effective remote teacher development.

Adam Robbins
Lead practitioner (whole school CPD), The Regis School
Managing Editor, CogSciSci

Abstract

The purpose of this research project was to understand the most effective ways of introducing to-be-learned material in teacher professional development delivered online. Prior research showed that preparing adults for learning using advance organisers (AOs) can improve learning outcomes (Mesmer-Magnus & Viswesvaran, 2010). However, whether AOs can also be used to facilitate learning in teacher professional development has yet to be determined. This research project was fully embedded in an online learning environment used to deliver teacher professional development. Participants (N = 134 teachers) were randomly divided into two groups. Each group experienced a different form of introductory material (an AO vs. a non-AO) presented ahead of to-be-learned material. A delayed learning outcome assessment was used to quantify the effects of the different types of introductory materials on learning after accounting for the effect of prior knowledge. Contrary to our hypothesis, we did not find an effect of type of introductory material on learning. This suggests that the theory underpinning AOs is not yet sufficiently precise to reliably guide the design of teacher professional development.

Key words: advance organiser, online learning, professional development, teachers

Do advance organisers help teachers learn new information during online professional development?

Advance organisers (AOs) are a diverse category of orienting introductory materials and activities, presented during the initial phase of instruction[1]. For example, an AO could come in the form of a pre-reading exercise that provides the learner with a scaffold to integrate to-be-learned material by activating related prior knowledge (Cannon-Bowers et al., 1998; Gurlitt, 2012; Jung et al., 2022). AOs are introduced before to-be-learned material to improve the learning thereof. AOs may include, but are not limited to, written text segments (Corkill et al., 1988), oral descriptions (Preiss & Gayle, 2006), examples (Gurlitt et al., 2012) or analogies (Gick & Holyoak, 1980), as well as graphical organisers (Robinson & Kiewra, 1995) like hierarchical charts (Potelle & Rouet, 2003) and concept maps (Gurlitt & Renkl, 2010).

AOs were initially developed in the 1960s (Ausubel, 1960, 1968). They were described as "introductory material at a high level of abstraction, generality, and inclusiveness" that deliberately introduce "relevant and appropriately inclusive subsuming concepts into cognitive structure" to provide "helpful ideational scaffolding which enhances the incorporability and longevity of the more detailed material in the learning passage" (Ausubel & Fitzgerald, 1961, p. 266). This definition has been criticised as vague (for instance, Barnes & Clawson, 1975). Within the assimilation encoding theory, Mayer (1979a, 1979b) suggested the following concrete characteristics for an AO: (1) comprises a short set of verbal or visual information; (2) is presented prior to the to-be-learned material; (3) unlike a summary, does not contain specific content from the to-be-learned material; (4) provides means of "generating" the logical relationships among the elements in the to-be-learned material, and (5) thereby influences the learner's *encoding* process.

Further, AOs should "be written in concrete fashion" and "contain an example that encourages the reader to note the *analogous relationship*" between the AO and to-be-learned material (Corkill, 1992, p. 63). AOs are hypothesised to fulfil three functions with respect to encoding processes (Cannon-Bowers et al., 1998; Mayer, 1979a): (1) guiding the learner's attention towards important aspects and relationships in the to-be-learned material; (2) providing a framework that can be used to organise the incoming information; and (3) activating relevant schema to facilitate the integration of new information into existing knowledge structures. Here, it is important that these functions would not be fulfilled if the AO was not presented, for example if the to-be-learned material itself already provides a suitable organisational framework to anchor incoming information. Through the activation of prior knowledge, AOs help to generate top-down semantic and structural overviews that can facilitate comprehension, thereby decreasing cognitive load during learning (Jung et al., 2022). The effect of AOs on learning is thought to be during information encoding rather than retrieval (Corkill, 1992).

In situations where the to-be-learned materials are completely unfamiliar, the main function of the AO is to provide initial scaffolding and facilitate learning by relating new material to existing knowledge (Ausubel & Fitzgerald, 1961; Mayer, 1979a, 1979b). These AOs are referred to as *expository* AOs. The condition of complete unfamiliarity is seldom met, however. As such, the to-be-learned material is often a variant of prior knowledge and the AO then additionally functions to "increase the discriminability of the learning passage from *analogous* and often conflicting ideas in the learner's cognitive structure" (Ausubel & Fitzgerald, 1961, p. 266). These AOs are

1. Ausubel (1968, p. 148) defined AOs initially as introductory materials presented at a "higher level of abstraction, generality and inclusiveness". However, current views extend beyond this (Corkill, 1992). This is reflected in the broad definition used here.

referred to as *comparative* AOs. While ideas consistent with the activated schema are thought to be assimilated, any inconsistent information is encoded as "discrete schema corrections" (Derry, 1984, p. 99). In cases where the learners possess relevant schemas, a comparative AO can help to highlight the relationship between existing knowledge and new information (Mayer, 1979a).

As such, comparative AOs share features with analogies. Analogies illustrate how two things are structurally alike. Hence, an analogy can help learners notice how to-be-learned material (the target) relates to something familiar (the source). For example, the operation of a city water supply can be used as an analogy for the cardiovascular system where the heart functions as the pump and arteries are parallel distribution pipes (Swain, 2000). Through analogies, knowledge is *transferred* by applying information or principles from the source to the target (Antonietti, 2012) – likely "by a process of mapping – finding a set of one-to-one correspondences (often incomplete) between aspects of one body of information and aspects of another" (Gick & Holyoak, 1983, p. 2). Hence, analogies – when presented ahead of to-be-learned materials – meet the characteristics of AOs proposed by Mayer (1979a, 1979b) by definition. More specifically, as analogies are used to transfer knowledge from the source to the target area, they do not contain specific information from the to-be-learned material. Likewise, the mapping process provides means of generating the logical relationships among the elements in the to-be-learned material. The mapping process further guides attention to relevant aspects of the to-learned material, provides a framework, and activates relevant schemas.

Examining the few AOs included in supplementary materials of previous studies shows that AOs are often analogies – even though the term analogy itself is not used. For instance, Ausubel's seminal Christianity-Buddhism comparative AO (Ausubel & Fitzgerald, 1961) was later described as "concrete with specific *analogies* for to-be-learned information and direct ties to knowledge that subjects already possessed" (Corkill et al., 1988, p. 77). Likewise, the AOs utilised by Corkill and colleagues (1988) point out the analogous relationships between a model car and models used in astronomy. In both studies, it was found that these AOs based on analogies increased knowledge retention of the to-be-learned materials (Ausubel & Fitzgerald, 1961; Corkill et al., 1988). Moreover, analogies facilitate problem solving (Gick & Holyoak, 1980) and induce schemas (Gick & Holyoak, 1983). Analogies are in fact so central to AOs that a review concluded that AOs should "contain an example that encourages the reader to note the analogous relationship" between the AO and to-be-learned materials (Corkill, 1992, p. 63).

These hypothesised benefits of AOs in general, and analogies in particular, make them potentially very useful for helping professionals learn new bodies of material as part of their training. For example, teachers may be learning about reading instruction, an area that involves specialist technical terms (morpheme, phoneme, etc.), understanding of numerous interacting subskills (vocabulary, fluency, etc.), and a sizable body of empirical evidence (for example, systematic synthetic phonics). Making sense of, and remembering, this large body of knowledge is challenging. AOs are a promising way of making this task more manageable for teachers, which would in turn be beneficial for their pupils. The increase in online professional development for teachers since the pandemic makes AOs – many of which are visual – particularly relevant to the learning design challenges currently faced by teacher educators (Bragg et al., 2021; Brown et al., 2021).

AOs were initially developed by Ausubel (1960, 1968) to test the hypothesis that the introduction of relevant subsuming concepts, that is, schemas, can enhance learning of new material. Since then, numerous research has evaluated the effectiveness of AOs in facilitating learning. An initial synopsis of 32 studies (Barnes & Clawson, 1975), which compared the number of studies that reported a significant effect of AOs against those that did not, indicated that there is no overall effect of AOs. However, the paper was methodologically flawed (Corkill, 1992; Luiten et al., 1980; Mayer, 1975). Thereafter, it has been argued that the effectiveness of AOs may depend on the conditions under which they are used (Mayer, 1979a, 1979b). These boundary conditions have been synthesised within assimilation encoding theory (Mayer, 1979a, 1979b). Early meta-analyses (Lott, 1983; Luiten et al., 1980; Stone, 1983) that integrated the effect sizes observed in individual studies reported small-to-moderate average effect sizes (that is, Glass' Delta;

Glass, 1977), indicating that AOs facilitate the retention of knowledge compared to control conditions. Importantly, observed variance in the integrated effect sizes suggest the effectiveness of AOs may vary depending on the design. For instance, Luiten and colleagues (1980) have shown that effects of AOs increase as the retrieval interval increases, up to a point where effects are nearly twice as large in delayed (more than 22 days) compared to immediate memory tests.

A more recent meta-analysis (Mesmer-Magnus & Viswesvaran, 2010) integrated effect sizes from studies ($k = 49$, $N = 4884$) where AOs were applied as part of a training programme (that is, professional development) involving adult trainees. They reported moderate-to-large effect sizes of AOs on learning, especially when combined with self-directed training methods. This aligns with proposals by others (Jung et al., 2022) that instructional models that rely on schema activation – a function of AOs – are particularly suited for self-paced online environments. While we are not aware of any quantitative, experimental evidence that evaluates the use of AOs in the professional development of teachers, qualitative data suggests that embedding AOs in teacher professional development may be a fruitful approach to improve the learning and understanding of new information (Chien, 2012). Based on these results we hypothesise that exposure to an AO (compared to a control condition) during teacher professional development ahead of new, to-be-learned material will facilitate learning of the new material. The aim of the current study was to experimentally test this hypothesis within a course that is part of a larger teacher professional development framework delivered online.

Methods

Participants

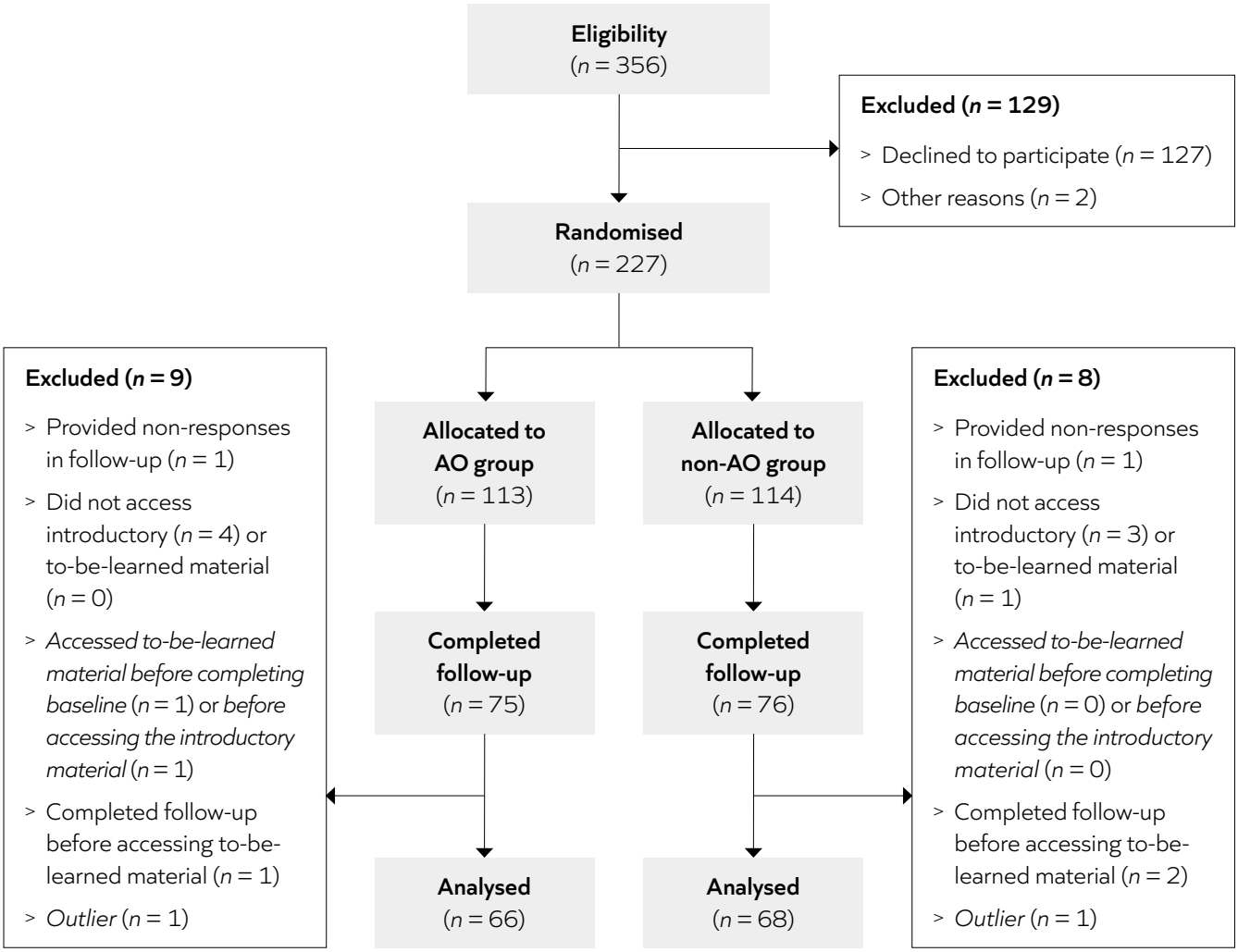
The sample was taken from a single cohort of fully qualified in-service teachers partaking in a teacher professional development programme, designed and delivered by Ambition Institute. The final sample of consenting participants should be considered a convenience sample. All 410 participants enrolled in the programme were invited to take part in the research project. Information about the research project was circulated together with other introductory information about the programme in an orientation course (see Figure S1). The orientation course preceded the programme and was available to all teachers enrolled in the programme; however, completion of the orientation course was not mandatory. The research project was further advertised in direct emails to enrolled programme participants as well as at in-person introductory events.

Given the large range of previously reported effect sizes, the aim was to recruit as many participants as possible and no stopping rule was applied. However, programme participants had to provide consent before 17 March 2023, so that research-related content was presented to consenting participants together with standard programme content released on 20 March 2023. On the cut-off date, a total of 356 participants had responded to the invite, out of which 229 programme participants had consented to take part in the research project. Two participants discontinued the programme, so only 227 participants were randomly allocated to one of two experimental conditions (see section on study design).

Data collection for the project closed on 31 October 2023, at 11:59pm while the programme was still ongoing. Up to this time point, we received 214 responses to the prior knowledge assessment and 151 responses to the learning outcome assessment. As such, we gained complete data (that is, pre- and post-test measures) from $N = 151$ participants. Note that participants who did not submit their responses to the learning outcome assessment did not withdraw from the programme or the research project; rather, they did not complete the assessment before data collection closed, most likely because they fell behind in the programme. As such, the attrition in the project is separate from the attrition of the programme. Upon careful inspection of the data, a total of 17 participants were excluded from the analyses[2]: two participants were identified who provided non-responses (for example, "dhbfgd") to at least 14 out of 15 short answer items included in the follow-up assessment; engagement data from 13 participants indicated that the experimental protocol was not followed and/or that the experimental manipulation was invalidated; and two participants were identified as outliers[3]. The participant flow is further illustrated in Figure 1.

2. Exclusion of these $n = 17$ participants did not impact the results (see Table 2).
3. The criteria for outlier removal were pre-registered. Any values that lie more than 1.5 interquartile ranges below the first or above the third quartile are considered outliers.

Figure 1. Consort flow chart



Note. The chart illustrates the participant flow and details reasons why participants were removed after the follow-up was completed but before data was analysed. Pre-registered reasons for exclusion are in italics.

After exclusions, the resulting sample size analysed was $N = 134$ participants. *A priori* power analysis indicated that a total sample size of 128 participants was required to detect a medium main effect of experimental group (that is, $f = 0.25$, numerator $df = 1$) after accounting for the effect of a single covariate in an Analysis-of-Covariance (ANCOVA) model with a power $1 - \beta = .80$ and type-1 error rate $\alpha = .05$.

In total, $n = 122$ participants voluntarily provided information on their age and gender. The mean reported age was 37.08 years ($SD = 8.48$). The sample consisted of 85.2% females. Demographic data for each group is shown in Table 1. The groups did not differ significantly in any of the characteristics (all $ps > .168$).

Table 1. Sample characteristics in each group

	AO group	Non-AO group	Group comparison
Gender (% female)	85.48%	85%	$\chi^2(1) = 0$, $p = 1$
Age	36.65 (8.94) [25; 56]	37.53 (8.03) [24; 53]	$t(119.35) = -0.578$, $p = .564$
Work experience	11.26 (7.39) [3; 30]	10.93 (7.07) [0; 27]	$t(119.98) = 0.248$, $p = .805$
Prior knowledge	5.8 (1.72) [2; 10]	6.22 (1.77) [2; 10]	$t(132) = -1.385$, $p = .168$

Note. For gender, age and work experience, only data from participants who provided the information voluntarily was used ($n = 122$, out of which 62 belonged to the AO group). For the prior knowledge assessment, data from the whole sample was used ($N = 134$). Table shows *M* (*SD*) [min; max]. Work experience was measured in years. The prior knowledge assessment had a maximum possible score of ten.

Study design

The research project was embedded within the professional development programme. The programme focused on effective teaching practices, including relevant background knowledge, and was tailored to leadership of literacy teaching within schools. The curriculum content was divided into six courses that lasted six weeks each. Each six-week course was further divided into three two-week modules, comprised of a "study" module in week A, and "paired practice apply" module in week B. During each week of a course, programme participants completed a module in an asynchronous, self-guided manner on Steplab (**steplab.co**). Steplab is a professional learning platform for schools and teachers and is used by Ambition Institute to deliver online learning materials using for example written articles or videos.

We used a between-subject design where participants were randomly assigned to one of two conditions. In the experimental group (AO group), participants were exposed to an AO before reading an evidence summary whereas participants in the control group (non-AO group) were exposed to a historical overview. We used block randomisation to randomly assign consenting participants to the experimental conditions to create groups with equal sample sizes (Altman & Bland, 1999). The randomisation was implemented using a custom-made R script. To evaluate the effectiveness of the experimental manipulation, participants' learning of the evidence summary was assessed in a delayed learning outcome assessment.

Additionally, participants' prior knowledge was quantified using a baseline prior knowledge assessment before the intervention. The study protocol was reviewed and approved by UCL Institute of Education Research Ethics Committee.

Materials and measurements

The study materials consist of (1) a prior knowledge assessment, (2) two introductory materials (the AO and historical overview), (3) the to-be-learned materials (an evidence summary on the development of reading abilities), and (4) a learning outcome assessment. Introductory materials as well as prior knowledge and learning outcome assessment were designed for the purpose of this study by an expert in the field of literacy.

To-be-learned materials

The to-be-learned material was a 5,156-word summary on scientific evidence related to the theory of how pupils learn to read. This evidence summary was part of the programme and was not adapted for research purposes. The evidence summary described how reading relies on two competencies — word recognition and language comprehension — and each of these competencies was explained in further detail. The readability of the evidence summary was estimated using an automated readability checker (**readabilityformulas.com/free-readability-formula-tests.php**), scoring the evidence summary to be at Grade Level 10 with a reading level of "fairly difficult to read".

Introductory materials

Based upon the literature reviewed above, an analogy was used as an AO. Analogies illustrate how two things are structurally alike. As such, an analogy can help learners notice how the to-be-learned material relates to something familiar. The analogy we developed (see supplementary materials) was based on the similarities between teaching someone to drive and teaching someone to read in so far that (1) both involve a complex, interacting array of knowledge and skills and (2) both are not capabilities that come naturally to most people. These points were elaborated upon in the context of teaching someone to drive before pointing out how this can be seen as akin to teaching reading (to a certain degree). The aim of the AO was to highlight the analogous relationship between the example of teaching someone to drive and the content of the to-be-learned material.

More specifically, to illustrate the first point, we explained that driving involves a complex interaction of skills and so does reading. In the same way a driving instructor needs to break down driving to help people learn, so a teacher needs to understand the individual components and break them down for the pupils learning to read. This was intended to help the programme participants organise and understand the array of components that were subsequently introduced in the to-be-learned materials, such as word recognition and language comprehension. To highlight the second point, we explained that neither driving nor reading are capabilities that come naturally to people as they constitute "biologically secondary knowledge" (Kalyuga, 2023). This was intended to help programme participants understand that opportunities for practice are necessary to learn these skills.

The AO was designed to meet the characteristics proposed by Mayer (1979a, 1979b). The resulting AO was (1) a short set of visual information in the form of written text; (2) presented before the to-be-learned material; (3) did not contain specific content included in the evidence summary, and (4) provided means of generating logical relationships among elements in the to-be-learned material by pointing out the similarities. As such, the AO was intended to guide the reader's attention, provide a framework for the incoming information, and activate relevant prior knowledge.

The non-AO discussed the development of literacy and societal progress as two mutually reinforcing processes. To match the AO in its overall structure, these two processes were mentioned and then later elaborated on. However, no connection was drawn to teaching reading. The non-AO was designed to match the AO with respect to criteria (1), (2), and (3), used a similar number of words, and was also presented ahead of the evidence summary. In contrast to the AO, however, the non-AO did not provide a framework that could be used to embed the to-be-learned material in relation to relevant prior knowledge.

Several steps were taken to ensure that the introductory materials were similar and only differed in respect to whether they activated relevant prior knowledge (for details, see supplementary methods and Table S1). The final versions of the introductory materials can be found in the supplementary materials.

Prior knowledge assessment

To assess participants' prior knowledge on the topic of the to-be-learned material (that is, development of reading abilities), a test containing ten single-choice items with four-alternative forced-choice response scales was developed by an expert in the area. For each item, participants had to select a single response that represented the correct answer according to their knowledge. The scope of the prior knowledge assessment was for each item to target a different construct relevant to the development of reading abilities (for example, "An orthography is…"). For each item, the correct answer (for example, "The set of conventions associated with a written language.") together with three plausible distractors (for example, "The set of conventions associated with a phonics system.") were presented. The order of presentation of the items and corresponding distractors was randomised, and the same random order was applied across all participants. All items were presented on a single screen. Additional information on the test development and piloting, together with the final items can be found in the supplementary methods and Table S2. The maximum possible score that could be obtained was ten that is, one point per correctly answered single-choice item.

Learning outcome assessment

To determine whether exposure to an AO facilitates learning, an expert in the field developed a learning outcome assessment for the study. The learning outcome assessment targeted information presented in the to-be-learned material and included items capturing knowledge retention and transfer. The learning outcome assessment consisted of 15 short answer items (for example, "Complete this sentence: The smallest chunk of meaning within a word is called a _____.") and eight multiple-choice items, each with four-alternative forced-choice response scales. For each presented multiple-choice item (for example, "Which of these are examples of comprehension strategies that might be used by pupils to support them to build meaning from a challenging text?"), participants were instructed to select all correct answers out of four alternatives they thought as correct (for example, "Clarifying", "Summarising", "Highlighting", "Blending").

For each of the short answer items (for example, "Please complete the following sentence: With extensive practice, the process of word recognition develops a sense of flow. This flow of words is called _____."), the correct solution was determined before the assessment was administered (that is, "(reading) fluency"). Additionally, all given answers were reviewed by a rater blind to the experimental condition and alternative acceptable solutions (for example, "prosody", "automaticity") were identified. A point was awarded if the participant's answer to an item contained any acceptable solution. For each multiple-choice item with its four answer alternatives, the pattern of selected/unselected alternatives was compared against the correct pattern. This means that a point was awarded if a correct alternative was correctly selected and if an incorrect alternative was correctly not selected. Therefore, there were four points available for each multiple-choice item. However, due to an error, one multiple-choice item was set up to be a single-choice item, that is, only one answer alternative could be selected. This item was scored as if it was a single-choice item and a point was awarded if the correct answer was selected, but not if the incorrect answers were not selected. The maximum total score on the multiple-choice items was therefore 7*4 + 1*1 = 29.

The learning outcome assessment together with additional information on development and piloting can be found in the supplementary information. The order of presentation of the items (and corresponding distractors, in case of multiple-choice items) was randomised, and the same random order was applied across all participants. All items were presented on a single screen. The maximum possible score that could be obtained when the 15 short answer items and eight multiple-choice items were combined was 44.

Learning experience questionnaire

To measure the subjective learning experience, we adapted the measures used by Wang and colleagues (2021). More specifically, we assessed learning satisfaction with the introductory and to-be-learned materials on a 7-point Likert scale (1 = "strongly agree" to 7 = "strongly disagree"); perceived difficulty of the to-be-learned material on a 7-point Likert scale (1 = "very, very easy" to 7 = "very, very difficult"); and mental effort in engaging with the to-be-learned materials on a 7-point Likert scale (1 = "very, very low mental effort" to 7 = "very, very high mental effort"). The items can be found in the supplementary materials.

Professional development programme and testing procedures

The research project was executed within Steplab. Content related to the research project was presented only to consenting participants. The prior knowledge assessment was embedded within the first module of Course 2, the AO-manipulated learning in the first module of Course 3, and the learning outcome assessment in the first module of Course 4. The release date of Course 2 was therefore considered as baseline timepoint (T_0). The data collection procedures are further illustrated in Figure 2 (see Figure S2 for screenshots). Within Steplab, courses are released sequentially, spaced six weeks apart. That means Course 3 was released six weeks after Course 2 ($T_1 = T_0 + 6$ weeks) and Course 4 was released another six weeks thereafter ($T_2 = T_1 + 6$ weeks $= T_0 + 12$ weeks). When a course was released, all three bi-weekly modules within a course could be completed by the programme participants in any order, though it was recommended to follow their numerical order. Programme participants were encouraged to complete the courses as they are released and to spend roughly an hour per week completing the weekly module. Within a module, content was presented in the form of objects, that is, a contained section on the webpage, that programme participants could "mark as completed" (see Figure S3 for screenshots).

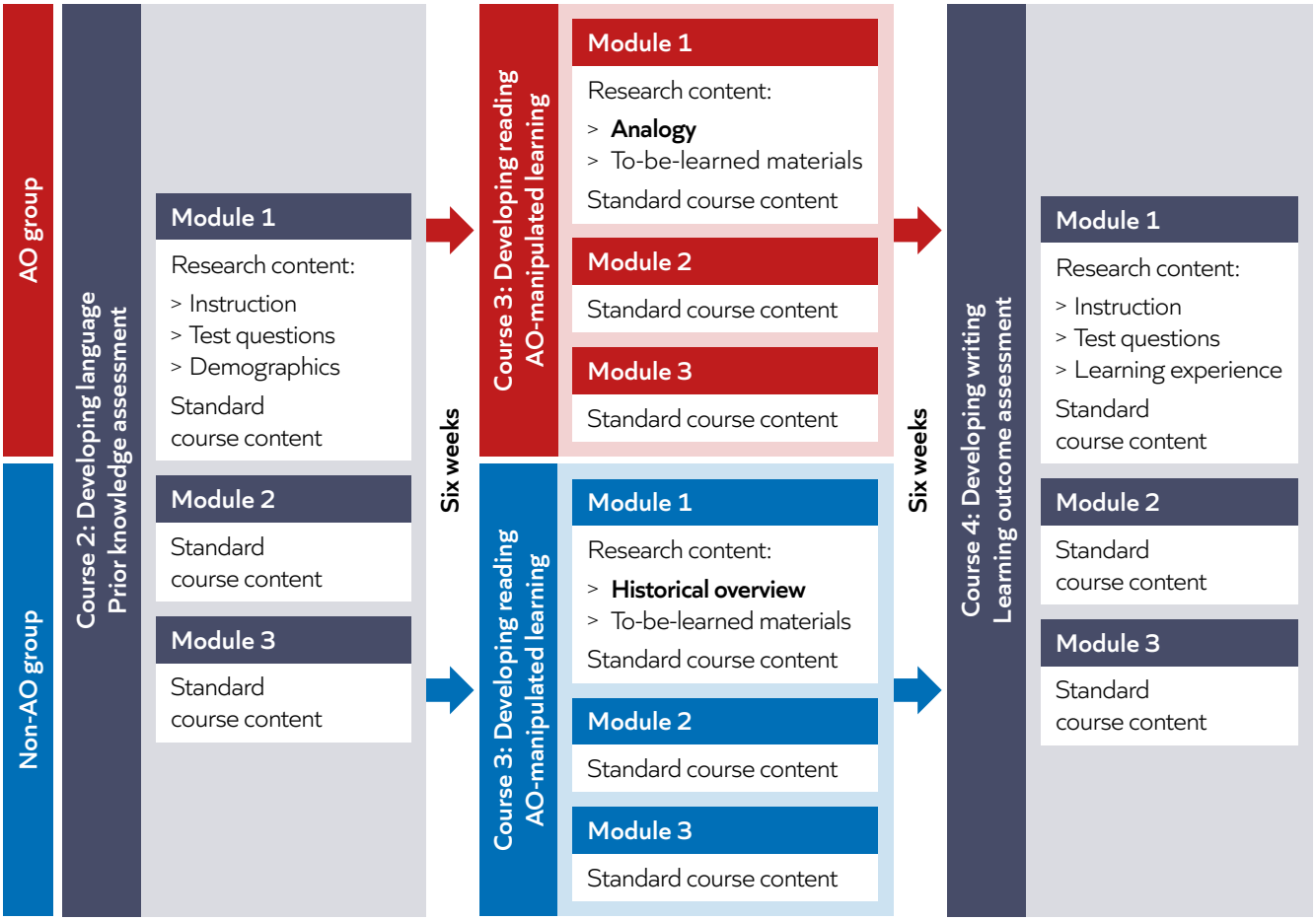
Research elements (the prior knowledge assessment, the AO-manipulated learning and the learning outcome assessment) were embedded as separate objects within the default module content of the programme and distinctly recognisable as such. The prior knowledge assessment consisted of instructions, the ten-item four-alternative forced-choice test and a demographic survey, all shown as separate objects. For the prior knowledge assessment, participants were instructed to select the answer alternative that they thought was correct and were encouraged to guess but discouraged from looking up the correct responses. Responses could only be submitted once each of the ten items had been completed. Upon completion of the prior knowledge assessment, participants were asked to voluntarily provide basic demographic information.

The AO manipulation consisted of the presentation of one of the introductory materials (AO or non-AO) immediately before the presentation of the to-be-learned material that was part of Module 1 of Course 3. The module started with a short video (same in both conditions), followed by the introductory materials (AO or non-AO) and the to-be-learned material (that is, evidence summary), all presented as separate objects on the same screen. Participants were instructed to carefully read the introductory material and indicate that they had done so by marking the section as complete. The learning outcome assessment consisted of instructions, 23 items and the learning experience survey. Again, all elements were independent objects. Participants were provided with matching instructions for each item type (short answer and multiple-choice items). Participants were further informed that incorrect answers in the multiple-choice items would be penalised. Further, they were instructed to not look up responses. Responses could only be submitted if each of the items had been completed. Upon completion of the learning outcome assessment, participants were asked to complete a short learning experience questionnaire.

The prior knowledge and learning outcome assessments were presented ahead of the course-specific content that was unrelated to the research project. The assessments were placed within courses that were not directly relevant to the AO-manipulated learning so that participants could not use the content presented in the to-be-learned material to complete the assessment without considerable effort, that is, navigating in between different courses with multiple clicks[4]. In case of prior knowledge, the assessment took place before the content of the AO manipulation (including the to-be-learned material) was released to participants.

4. As in any study that collects assessments online without an experimenter present, there is a risk that participants use online resources to complete the assessment. Given the low stakes and explicit instruction to not look up the correct responses, this risk can be deemed neglectable in the context of this research project.

Figure 2. Illustration of data collection



Note. Each programme course lasted for six weeks, and courses were released automatically to participants. Each course consisted of three modules, where each module had a study and apply component. While modules could be completed in any order, programme participants were encouraged to complete modules sequentially. Research content was embedded within the study component of the first module of the respective courses. Text boxes show the research content in the order in which it was presented to participants. Content that differed between the experimental groups is highlighted in bold. The illustration only contains programme courses relevant to the research project. For details on the content of the courses and modules, please refer to Figure S2, that contains screenshots from the Steplab environment.

Data analysis

Data were collected using Steplab. Steplab data were accessible through a PostgreSQL database maintained by Aircury (**www.aircury.com/**) and were queried using custom scripts. All analyses were pre-registered (**osf.io/vfwsc/**) unless indicated otherwise, and conducted using R version 4.3.2 (R Core Team, 2023). While data cannot be shared, all code used during this project is available on Github (**github.com/stefaniemeliss/AO_Project**).

Our primary hypothesis stated that the exposure to an AO would improve the encoding of subsequently presented, to-be-learned material. We were interested in the main effect of experimental condition, that is, whether an AO or historical overview, presented as introductory material, affects learning of the evidence summary (measured as the sum score of the learning outcome assessment), after controlling for prior knowledge at baseline (measured as the sum score of the prior knowledge assessment). The sum scores were standardised, and the independent variable was dummy-coded (0 = "non-AO", 1 = "AO"). To test this hypothesis, an analysis-of-covariance (ANCOVA) was used. The ANCOVA approach is appropriate for analysing data from two-group pre-post studies (Dugard & Todman, 1995). Effects were quantified using η_p^2 (partial Eta squared) as effect size.

An ANCOVA can be seen as a combination of analysis-of-variance (ANOVA) and regression. More specifically, the dependent variable is regressed on the covariate(s) first and the residuals of the model (that is, unexplained variance) are subject to a between-subject factorial ANOVA with two levels. The ANOVA determines whether there is an overall effect of group by comparing the variance observed within each group to the variance observed between the groups. If the latter is larger, this suggests that there is a significant main effect of group. By regressing out the covariate(s), the within-group error variance is reduced, thereby increasing the power to find a significant effect of group.

Results

Primary hypothesis testing

After checking the assumptions (see supplementary methods), an ANCOVA was run to determine the effect of introductory material on the learning outcome assessment score after controlling for prior knowledge assessment score. The results of the ANCOVA (see Table 2) showed that after adjustment for prior knowledge assessment score, there was no statistically significant difference in learning outcome assessment score between the groups, $F(1, 131) = 1.43$, $p = .234$. The covariate prior knowledge significantly predicted performance in the learning outcome assessment, $F(1, 131) = 12.33$, $p = .001$[5]. The means of covariate and outcome measure are shown in Table 3 and Figure 3.

Table 2. ANCOVA results

	SS	*df*	MS	*F* value	*p* value	η_p^2
Primary hypothesis						
AO effect	1.3	1	1.3	1.43	.234	0.01
Prior knowledge	11.24	1	11.24	12.33	.001	0.09
Residuals	119.35	131	0.91			

Exploratory analysis

No exclusions						
AO effect	0.03	1	0.03	0.03	.858	0
Prior knowledge	17.39	1	17.39	19.45	< .001	0.12
Residuals	132.32	148	0.89			
Pre-registered exclusions only						
AO effect	1.2	1	1.2	1.3	.256	0.01
Prior knowledge	10.51	1	10.51	11.44	.001	0.08
Residuals	123.11	134	0.92			
Release date						
AO effect	0.78	1	0.78	0.85	.358	0.01
Prior knowledge	10.37	1	10.37	11.28	.001	0.08
Residuals	115.83	126	0.92			
Non-sequential module completion						
AO effect	1.83	1	1.83	1.96	.165	0.02
Prior knowledge	9.66	1	9.66	10.34	.002	0.08
Residuals	110.25	118	0.93			

5. The different scaling of prior knowledge (maximum possible score = 10) and learning outcome assessment (maximum possible score = 44) prevented us from being able to determine whether overall learning occurred. Performance in prior knowledge and learning outcome assessment were significantly correlated, $r = .30$, $p < .001$.

	SS	*df*	MS	*F* value	*p* value	η_p^2
Time intervals						
AO effect	0.78	1	0.78	0.82	.366	0.01
Prior knowledge	10.12	1	10.12	10.72	.001	0.08
Interval 1	1.23	1	1.23	1.3	.256	0.01
Interval 2	0	1	0	0	.963	0
Residuals	109.45	116	0.94			
Engagement duration						
AO effect	0.87	1	0.87	0.97	.326	0.01
Prior knowledge	7.34	1	7.34	8.19	.005	0.07
Duration	1.88	1	1.88	2.1	.150	0.02
Residuals	100.4	112	0.9			
Demographics						
AO effect	2.62	1	2.62	2.84	.094	0.02
Prior knowledge	9.91	1	9.91	10.77	.001	0.08
Gender	1.73	1	1.73	1.88	.173	0.02
Experience	0.08	1	0.08	0.09	.763	0
AO x experience	1.05	1	1.05	1.14	.287	0.01
Residuals	106.79	116	0.92			
Prior knowledge below median						
AO effect	2.73	1	2.73	3.19	.080	0.06
Prior knowledge	3.32	1	3.32	3.88	.054	0.07
Residuals	41.94	49	0.86			

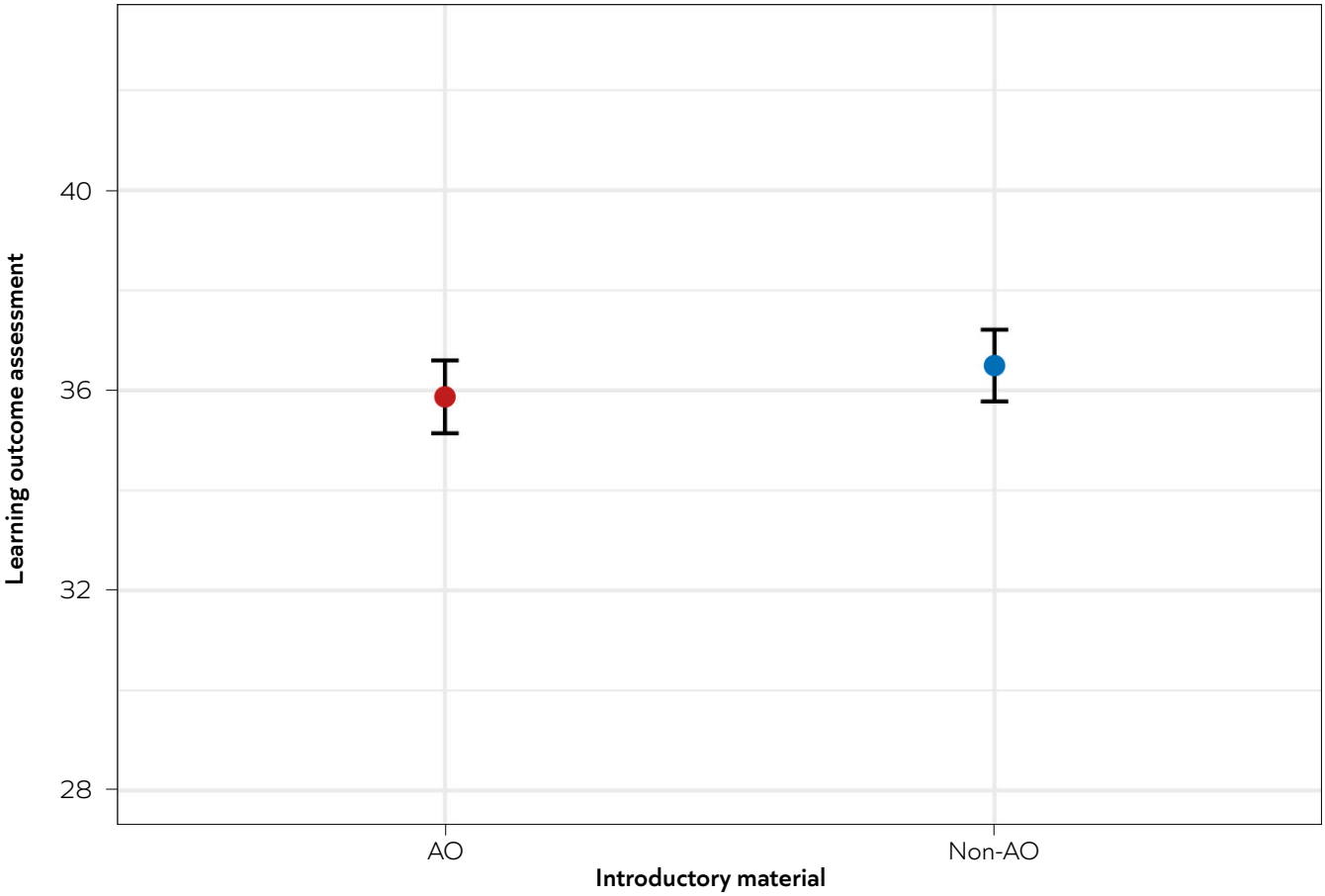
Note. The table summarises the results of all ANCOVA models run. SS = sum of squares, *df* = degrees of freedom, MS = mean square, η_p^2 = partial Eta squared.

Table 3. ANCOVA adjusted means

Group	Prior knowledge (covariate)	Observed learning outcome	Adjusted learning outcome
AO	5.8 (0.21)	35.76 (0.39)	35.87 (0.37)
Non-AO	6.22 (0.21)	36.6 (0.37)	36.49 (0.37)

Note. Table shows *M* (*SE*) of the covariate and outcome measure as well as the covariate-adjusted outcome measure separately for each group.

Figure 3. Adjusted means of the learning outcome assessment



Note. The learning outcome assessment scores plotted here were adjusted for performance in the prior knowledge assessment. Error bars represent 95% confidence intervals

Exploratory analysis

As explained earlier and illustrated in Figure 1, some participants were excluded from the analysis for various reasons. We repeated the analysis with data from all $N = 151$ participants for whom prior knowledge and learning outcome assessment data were available. This showed that the exclusion of participants did not impact the results (see Table 2).

There may have been cases where the participants had fallen behind in their online studies and Course 3 (containing the AO manipulation) was released before they had completed the prior knowledge assessment in Course 2, Module 1. Course 3 was released to participants on 22 May 2023, and some participants in the experimental ($n = 3$) and control ($n = 2$) groups completed the prior knowledge assessment after this date. As shown in Table 2, their exclusion did not impact the results.

Participants may have completed modules within Course 3 in a non-sequential order. In these cases, sensitivity analyses were conducted to determine whether the inclusion or exclusion of such participants impacted the results. To determine whether participants completed the modules in Course 3 in sequential order, timestamps of when the modules were marked as completed were used. If timestamps were missing for one or more

modules, this was regarded as equivalent to completing modules in a non-sequential order. Some participants in the experimental (n = 7) and control (n = 6) group had completed the modules in Course 3 in a non-sequential order. The respective participants were excluded before repeating the same ANCOVA analysis reported above. Again, the exclusion of these participants did not impact the results (see Table 2).

The time intervals between the initial engagement with (1) introductory material and evidence summary and (2) evidence summary and learning outcome assessment can be seen as proxies for the respective retention intervals. We were hence interested to determine whether the retention intervals influenced the results. The intervals were calculated using the timestamps associated with completion of the objects. The intervals between accessing the introductory material and evidence summary ranged from 4 to 4,830,053 seconds. The intervals between accessing the evidence summary and completing the learning outcome assessment ranged from 0.021 to 145.821 days. The intervals were checked for outliers to identify values that lie more than 1.5 interquartile ranges (IQR) below the first or above the third quartile. With respect to the first interval, n = 13 outliers were identified using the same IQR criteria described above. For the second interval, there were n = 0 outliers. For the analysis, all outliers were removed. The average time interval between accessing the introductory material and evidence summary was 93.96 seconds (SD = 63.74). The average interval between accessing the to-be-learned material and completing the learning outcome assessment was 86.57 days (SD = 34.51). The time intervals were standardised and included as covariates. The inclusion of time intervals did not impact the results (see Table 2).

The time in between participants first accessing the object containing the introductory materials and marking the object as completed can be seen as a proxy for how long they spent engaging with the introductory material. As such, we were interested to determine whether the inclusion of the engagement duration as a covariate impacts the results. This analysis was not pre-registered. The interval between accessing the introductory material and evidence summary ranged from 2 to 8,791,465 seconds. There were n = 18 outliers identified using the IQR criteria. These outliers were removed before conducting the analysis. The average engagement duration was 94.24 seconds (SD = 68.65). The duration was standardised and included as a covariate. However, the results were not impacted by this (see Table 2).

For the subset of participants that voluntarily provided demographic information, gender and experience as a classroom teacher were included as additional covariates to test whether including additional demographic variables impacted the results. Age was not included due to being highly correlated with experience, r = .76, p < .001, and experience was standardised. For this model, the independent variable was effect-coded (-1 = "non-AO", 1 = "AO") and a term modelling the interaction between experience and AO exposure was included. However, as shown in Table 2, controlling for demographic variables did not influence the results.

The theory underlying AOs suggests that AOs are particularly beneficial if learners are inexperienced. We hence repeated the ANCOVA analysis in a subset of participants who scored below the sample median in the prior knowledge assessment. This analysis was not pre-registered. N = 52 participants were identified whose performance in the prior knowledge assessment was below the sample median of 6. Using the subset of participants, we found a non-significant difference in the learning outcome assessment score between both groups after adjusting for prior knowledge, $F(1, 49) = 3.19$, p = .080. There was also a non-significant predictive effect of the covariate prior knowledge on the performance in the learning outcome assessment, $F(1, 49) = 3.88$, p = .054. The descriptive data suggests that the direction of the non-significant group effect was opposite to what we hypothesised. The performance in the learning outcome assessment adjusted for prior knowledge was higher in the non-AO group ($M_{adjusted}$ = 35.94, SE = 0.62) compared to the AO group ($M_{adjusted}$ = 34.48, SE = 0.53).

Learning experience questionnaire

To assess satisfaction with the introductory materials, participants were asked to rate the item "Overall, I was satisfied with the introductory material presented ahead of the evidence summary in Module 1 of Course 3 ('Learning to read')." on a 7-point Likert scale from 1 = "Strongly disagree" to 7 = "Strongly agree". No significant difference in satisfaction with introductory material was found between the AO (M = 5.35, SD = 1.27) and non-AO (M = 5.4, SD = 1.24) group, t(131.55) = -0.22, p = .823.

To assess satisfaction with to-be-learned material, participants were asked to rate the item "Overall, I was satisfied with the evidence summary presented in Module 1 of Course 3 ('Learning to read')." on a 7-point Likert scale from 1 = "Strongly disagree" to 7 = "Strongly agree". No significant difference in satisfaction with to-be-learned material was found between the AO (M = 5.47, SD = 1.08) and non-AO (M = 5.49, SD = 1.22) group, t(131.08) = -0.08, p = .938.

To assess effort, participants were asked to rate the item "How much mental effort did you invest to learn the content from the introductory material and evidence summary presented in Module 1 of Course 3 ('Learning to read')?" on a 7-point Likert scale from 1 = "Very, very low mental effort" to 7 = "Very, very high mental effort". No significant difference in effort was found between the AO (M = 4.91, SD = 0.7) and non-AO (M = 4.79, SD = 1) group, t(119.72) = 0.77, p = .440.

To assess difficulty, participants were asked to rate the item "How difficult was it for you to learn the content from the introductory material and evidence summary presented in Module 1 of Course 3 ('Learning to read')?" on a 7-point Likert scale from 1 = "Very, very easy" to 7 = "Very, very difficult". No significant difference in difficulty was found between the AO (M = 4.27, SD = 0.67) and non-AO (M = 4.22, SD = 0.75) group, t(131.08) = 0.42, p = .672.

Discussion

The aim of this study was to determine whether an AO could be used within the context of professional development for teachers delivered online. We hypothesised that the presentation of an AO (compared to a historical overview) would facilitate the encoding of to-be-learned materials presented thereafter. Contrary to our hypothesis, we did not find any evidence to suggest that AOs have a beneficial effect in the context of teacher professional development delivered online, both in terms of cognitive learning as well as the subjective learning experience.

Our results are inconsistent with meta-analytic evidence suggesting that AOs lead to learning gains when administered to adult participants in the context of a training programme (Mesmer-Magnus & Viswesvaran, 2010). The meta-analysis found positive effects of AOs in graphic and textual forms and effects were more pronounced in self-directed compared to traditional learning. This suggests that while the use of AOs may be beneficial in some training contexts, the effects do not generalise to all professions and/or professional development. This is in line with conclusions of a previous meta-analysis on oral AOs (Preiss & Gayle, 2006).

It is also important to note that the prior meta-analysis (Mesmer-Magnus & Viswesvaran, 2010) was conducted before the start of the open science movement in psychology. It is therefore likely that most, if not all, original studies were not pre-registered[6]. Publications without pre-registrations tend to find larger effect sizes compared to pre-registered studies (Schäfer & Schwarz, 2019). Assuming this to be the case, the results of the meta-analysis may have over-estimated the true effect size. While our study was adequately powered to detect a medium effect, we cannot exclude that the effect in an applied setting may be smaller than suggested by the meta-analysis. Researchers and practitioners will have to determine whether this may be relevant to them on a case-by-case basis.

It has been proposed that AOs facilitate learning and retention of unfamiliar materials. As the theory developed further, more specific predictions about the effectiveness of AOs were made. For instance, it has been proposed that AOs work best when the to-be-learned material is poorly organised, the material is unfamiliar, and when subjects are inexperienced (Mayer, 1979b). The latter has also been referred to as "*treatment x knowledge*" or *"treatment x ability interaction*". The participants in our study were qualified teachers with on average over a decade of experience who were enrolled in a professional development programme directed towards leadership in literacy. As such, the participants may have been too experienced (yet, see the median performance in the prior knowledge assessment was roughly 60%). However, when modelling the interaction terms between AO exposure and prior knowledge[7] (see Table S3) or experience, respectively, the interaction terms did not predict learning either. This is in line with the results of early meta-analyses on the effectiveness of AOs that failed to observe the proposed interactions (Luiten et al., 1980; Stone, 1983). Likewise, the to-be-learned material used here was well-organised and of high quality as it was taken from the standard content of the programme. Some studies have reported that exposure to an AO decreases learning in situations in which the to-be-learned materials were already well-organised (Grotelueschen & Sjogren, 1968; Lesh, 1976; Mayer, 1978; Schumacher et al., 1975). Taken together, the level of experience of our participants and the quality of to-be-learned materials may explain the null results reported here. Nevertheless, when we repeated the analysis for the subset of participants that had comparatively less prior knowledge, we again did not find that AOs facilitate learning. This suggests that the level of experience of the teachers in our sample cannot fully account for the lack of findings.

6. Our attempts to retrieve a comprehensive list of studies included in the meta-analysis were unsuccessful. We were hence not able to test this claim.
7. The correlation between prior knowledge and learning outcome assessment was only moderate. However, this may have been caused – at least partly – by different formats utilised in the assessments.

There is still debate about what makes a good AO (Corkill, 1992). For instance, while Ausubel's original definition (1968) suggested that AOs should be abstract, later research found that concrete AOs outperform abstract AOs (Corkill et al., 1988). While various attempts have been made to define AOs, definitions often lack specificity and exemplification to guide application. For instance, the AOs used in Ausubel's early work (for example, Ausubel, 1960; Ausubel & Fitzgerald, 1961; Ausubel & Youssef, 1963) were not included as appendices in the articles. This makes it difficult to conduct research on the effectiveness of AOs or to use them as instructional devices. As discussed earlier, the AO used here was designed specifically for the purpose of this study based on characteristics proposed by Mayer (1979a, 1979b). We further aimed to create the "optimum conditions" for AOs (Corkill, 1992, p. 65): (1) learners will not make appropriate connections between prior knowledge and the to-be-learned material without the help of the AO; (2) learners attend to the AO; (3) learners study introductory and to-be-learned materials for appropriate lengths of time; and (4) learning outcomes are assessed using delayed recall. All these conditions were addressed in our study[8].

To design the AO, that is, introductory materials presented at a "higher level of abstraction, generality and inclusiveness" (Ausubel, 1968, p. 148), we decided that the analogy of teaching someone to learn how to drive would be an effective anchor for the to-be-learned materials related to teaching pupils how to read. While we perceived this to be a suitable level of abstraction, generality and inclusiveness, some may argue that it is too abstract to activate relevant prior knowledge to help encode specific technical linguistic knowledge such as definitions of morphemes and phonemes. This tension illustrates the ambiguity around how to operationalise AOs to guide instruction and learning in a real-world setting. This would hence mean that the concept of AOs would need to be refined and better exemplified to be useful in the context of instruction science in general and teacher education in particular. This is further complicated by the fact that AOs are "relative to the particular learner and subject matter" (Mayer, 1979a, p. 382), making a one-size-fits-all approach difficult to design.

A possible limitation of the current study is that participants in the programme studied introductory and to-be-learned materials in a self-directed manner. While the programme provides participants with adequate study time, we could not ensure that participants indeed carefully attended to the AO. However, our proxy for the duration of engagement with the introductory materials ($M = 94.24$ seconds, $SD = 68.65$) suggests that participants indeed attended to the materials. The average engagement time overlaps with estimates of how long it would take to silently read circa 350 English words, assuming 238 non-fiction words per minute as reading speed (Brysbaert, 2019). Likewise, self-directedness is an inherent feature of online learning, and the aim of this research was to test the effectiveness of AOs in applied settings. A related concern here refers to the retention interval between the presentation of the introductory and to-be-learned materials and the learning outcome assessment. The learning outcome assessment was embedded in a course that was released six weeks after the course containing the introductory and to-be-learned materials. However, the course release dates often differed from the date when a participant engaged with the content. This is indeed reflected in the observed time intervals between accessing the to-be-learned materials and completing the learning outcome assessment ($M = 86.57$ days, $SD = 34.51$) that are closer to 12 weeks. It is possible that the AO created an initial memory trace that did not persist for this duration; indeed, the length of the interval provides ample opportunity for forgetting. However, in the context of professional development, the aim often is to acquire knowledge and skills that can readily be used and applied over long timespans. Thus, while AOs may be beneficial in some contexts, based on the results reported here, they do not seem to provide an answer to the learning design challenges currently faced by teacher educators (Bragg et al., 2021; Brown et al., 2021).

To conclude, in one of the first pre-registered experimental tests in this literature, we find that adding an AO to online professional development for experienced teachers does not improve learning. This suggests that the underpinning theory is not yet specific enough to reliably guide the use of AOs in such settings. In the absence of better theoretical and practical guidance on how to operationalise AOs, teacher educators may struggle to make productive use of AOs.

8. Please note that conditions (1) are (2) can only indirectly be demonstrated. Numerically speaking, the correlation between prior knowledge and learning outcome assessment were larger in the AO ($r = .40$, $p < .001$) compared to the non-AO group ($r = .19$, $p = .118$). Likewise, the engagement duration (discussed below) suggests that learners at least partially attended the AO.

References

Altman, D. G., & Bland, J. M. (1999). How to randomise. *BMJ: British Medical Journal, 319(*7211), 703–704.

Antonietti, A. (2012). Analogy/Analogies: Structure and Process. In N. M. Seel (Ed.), *Encyclopedia of the Sciences of Learning* (pp. 234–235). Springer US.

Ausubel, D. P. (1960). The use of advance organizers in the learning and retention of meaningful verbal material. *Journal of Educational Psychology, 51*(5), 267–272.

Ausubel, D. P. (1968). *Educational Psychology: A cognitive View.* Winston.

Ausubel, D. P., & Fitzgerald, D. (1961). The role of discriminability in meaningful learning and retention. *Journal of Educational Psychology, 52*(5), 266–274.

Ausubel, D. P., & Youssef, M. (1963). Role of discriminability in meaningful paralleled learning. *Journal of Educational Psychology, 54*, 331–336.

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39*(3), 445–459.

Barnes, B. R., & Clawson, E. U. (1975). Do Advance Organizers Facilitate Learning? Recommendations for Further Research Based on an Analysis of 32 Studies. *Review of Educational Research, 45*(4), 637–659.

Bragg, L. A., Walsh, C., & Heyeres, M. (2021). Successful design and delivery of online professional development for teachers: A systematic review of the literature. *Computers & Education, 166,* 104158.

Brown, C., Correll, P., & Stormer, K. J. (2021). The "new" normal: Re-imagining professional development amidst the COVID-19 pandemic. *Middle School Journal, 52*(5), 5–13.

Brysbaert, M. (2019). How many words do we read per minute? A review and meta-analysis of reading rate. *Journal of Memory and Language, 109,* 104047.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods, 46*(3), 904–911.

Cannon-Bowers, J. A., Rhodenizer, L., Salas, E., & Bowers, C. A. (1998). A Framework for Understanding Pre-Practice Conditions and Their Impact on Learning. *Personnel Psychology, 51(*2), 291–320.

Chien, C.-W. (2012). Use of Graphic Organizers in a Language Teachers' Professional Development. *English Language Teaching, 5*(10), 49–57.

Corkill, A. J. (1992). Advance organizers: Facilitators of recall. *Educational Psychology Review, 4,* 33–67.

Corkill, A. J., Bruning, R. H., & Glover, J. A. (1988). Advance Organizers: Concrete versus Abstract. *The Journal of Educational Research, 82*(2), 76–81.

Derry, S. J. (1984). Effects of an organizer on memory for prose. *Journal of Educational Psychology, 76*, 98–107.

Dugard, P., & Todman, J. (1995). Analysis of Pre-test-Post-test Control Group Designs in Educational Research. *Educational Psychology, 15(*2), 181–198.

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology, 12*(3), 306–355.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15*(1), 1–38.

Glass, G. V. (1977). 9: Integrating Findings: The Meta-Analysis of Research. *Review of Research in Education, 5*(1), 351–379.

Grotelueschen, A., & Sjogren, D. D. (1968). Effects of differentially structured introductory materials and learning tasks on learning and transfer. *American Educational Research Journal, 5*(2), 191–202.

Gurlitt, J. (2012). Advance Organizer. In N. M. Seel (Ed.), *Encyclopedia of the Sciences of Learning* (pp. 148–151). Springer US.

Gurlitt, J., Dummel, S., Schuster, S., & Nückles, M. (2012). Differently structured advance organizers lead to different initial schemata and learning outcomes. *Instructional Science, 40*(2), 351–369.

Gurlitt, J., & Renkl, A. (2010). Prior knowledge activation: How different concept mapping tasks lead to substantial differences in cognitive processes, learning outcomes, and perceived self-efficacy. *Instructional Science, 38*(4), 417–433.

Jung, E., Lim, R., & Kim, D. (2022). A Schema-Based Instructional Design Model for Self-Paced Learning Environments. *Education Sciences, 12*(4), 271.

Kalyuga, S. (2023). Evolutionary Perspective on Human Cognitive Architecture in Cognitive Load Theory: A Dynamic, Emerging Principle Approach. *Educational Psychology Review, 35*(3), 91.

Lesh, R. A. (1976). The influence of two types of advanced organizers on an instructional unit about finite groups. *Journal for Research in Mathematics Education, 7*(2), 87–91.

Lott, G. W. (1983). The effect of inquiry teaching and advance organizers upon student outcomes in science education. *Journal of Research in Science Teaching, 20*(5), 437–451.

Luiten, J., Ames, W., & Ackerson, G. (1980). A Meta-analysis of the Effects of Advance Organizers on Learning and Retention. *American Educational Research Journal, 17*(2), 211–218.

Mayer, R. E. (1975). Information processing variables in learning to solve problems. *Review of Educational Research, 45*(4), 525–541.

Mayer, R. E. (1978). Advance organizers that compensate for the organization of text. *Journal of Educational Psychology, 70*(6), 880–886.

Mayer, R. E. (1979a). Can Advance Organizers Influence Meaningful Learning? *Review of Educational Research, 49*(2), 371–383.

Mayer, R. E. (1979b). Twenty years of research on advance organizers: Assimilation theory is still the best predictor of results. I*nstructional Science, 8*(2), 133–167.

Mesmer-Magnus, J., & Viswesvaran, C. (2010). The role of pre-training interventions in learning: A meta-analysis and integrative review. *Human Resource Management Review, 20*(4), 261–282.

Potelle, H., & Rouet, J.-F. (2003). Effects of content representation and readers' prior knowledge on the comprehension of hypertext. *International Journal of Human-Computer Studies, 58*(3), 327–345.

Preiss, R. W., & Gayle, B. M. (2006). A Meta-Analysis of the Educational Benefits of Employing Advanced Organizers. In B. M. Gayle, R. W. Preiss, N. Burrell, & M. Allen (Eds.), *Classroom communication and instructional processes: Advances through meta-analysis* (pp. 329–344). Lawrence Erlbaum Associates Publishers.

R Core Team. (2023). R: *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. **www.R-project.org/**

Robinson, D. H., & Kiewra, K. A. (1995). Visual argument: Graphic organizers are superior to outlines in improving learning from text. *Journal of Educational Psychology, 87*, 455–467.

Schäfer, T., & Schwarz, M. A. (2019). The Meaningfulness of Effect Sizes in Psychological Research: Differences Between Sub-Disciplines and the Impact of Potential Biases. *Frontiers in Psychology, 10,* 813.

Schumacher, G. M., Liebert, D., & Fass, W. (1975). Textual organization, advance organizers and the retention of prose material. *Journal of Reading Behavior, 7*(2), 173–180.

Stone, C. L. (1983). A Meta-Analysis of Advance Organizer Studies. *The Journal of Experimental Education, 51*(4), 194–199.

Swain, D. P. (2000). The water-tower analogy of the cardiovascular system. *Advances in Physiology Education, 24*(1), 43–50.

Wang, X., Mayer, R. E., Zhou, P., & Lin, L. (2021). Benefits of interactive graphic organizers in online learning: Evidence for generative learning theory. *Journal of Educational Psychology, 113,* 1024–1037.

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods, 45*(4), 1191–1207.

Supplementary information

Supplementary methods

Linguistic equality of introductory materials

The automatic readability checker was used to verify that the AO had the same readability grade level as the to-be-learned material. This showed that the to-be-learned material and AO are both "fairly difficult to read" (grade level 10, reading age 14-15 years). Next, the automatic readability checker was applied to the non-AO to iteratively modify the non-AO to match the AO. After a satisfactory match between the introductory materials was achieved, the lexical characteristics for the final versions were generated using the English Lexicon Project database (Balota et al., 2007), available at **elexicon.wustl.edu/index.html**.

Results of the direct comparison between AO and non-AO using the automatic readability checker and averaged data from the English Lexicon Project can be found in Table S1. AO and non-AO have equivalent readability consensus scores. The word statistics (for example, word count, average word/sentence length), (psycho-) linguistic characteristics (for example, word frequency, age of acquisition, average standardised reaction time in a lexical decision task) as well as semantic characteristics (for example, concreteness or emotional valence) were highly similar. The only noticeable difference between AO and non-AO emerges in the ratio of unique to repeated words (41:59 in the AO, 50:50 in the non-AO). However, we argue that a higher proportion of repeated words in the AO is likely an inherent characteristic of an analogy where two things are compared with one another to highlight correspondence and similarity.

Piloting of prior knowledge assessment

After initial development, the items were piloted in an independent sample of $N = 111$ teachers to determine whether each item induces sufficient variance in the response (see **osf.io/vfwsc/** for details). Analysing the responses indicated that performance was significantly above chance level (that is, 25%) for each of the ten items, all correct response rates ≥ 40.54%, all $ps \leq .0002$, surviving stringent Bonferroni correction of $\alpha / k = 0.05 / 10 = 0.005$. However, to increase difficulty and variance, items and/or distractors were modified if the correct response rate in the pilot study was ≥ 80%. For items with high correct response rates, we tried to increase the difficulty by making the distractors more plausible, for example, by changing "A word" to "The sound of a whole word" as a plausible distractor for the item "A grapheme is an individual letter or a group of letters that represents…". In cases where this was not possible, the item itself was modified without affecting the scope of the item and thereby the breadth of the prior knowledge assessment. For instance, the original item "Pupils make inferences when they read…" was changed to "In terms of reading, inferences are conclusions that…", as well as all distractors. The wording of items was also modified in response to participant feedback to avoid ambiguity. For instance, the item "Which of these comprehension strategies does not have a sound basis in evidence?" was reworded to "Which of these comprehension strategies has the least sound basis in evidence?". The resulting prior knowledge test can be found in Table S2. The order of presentation of the ten items and corresponding distractors was randomised, and the same random order was applied across all participants. All items were presented on a single screen.

Piloting of learning outcome assessment

We initially created a set of 23 items capturing knowledge retention and transfer. The resulting learning outcome assessment consisted of eight short answer items (for example, "Complete this sentence: The smallest chunk of meaning within a word is called a _____.") and 15 multiple-choice items, each with four-alternative forced-choice response scales. For each presented item (for example, "Which of these are examples of comprehension strategies that might be used by pupils to support them to build meaning from a challenging text?"), participants were instructed to select all correct answers out of four alternatives they thought as correct (for example, "Clarifying", "Summarising", "Highlighting", "Blending").

The items were piloted in an independent sample of $N = 116$ teachers who were randomly assigned to read one of the two introductory materials (that is, the AO or non-AO) before completing the learning outcome assessment (see **osf.io/vfwsc/** for details). Teachers did not see the to-be-learned materials used in the main study. The purpose of this pilot study was to ensure that the content of the introductory materials alone would not impact the performance of the learning outcome assessment. To score the performance of the pilot study, the same strategy as outlined in the main text was applied. As such, each short answer item could be awarded with a point and for each multiple-choice item, the maximum score was 4.

Analyses showed that performance in the learning outcome assessment did not differ significantly between both groups, $t(110.07) = 0.88$, $p = .382$. This suggests that exposure to the introductory material itself without the to-be-learned material did not facilitate performance in the learning outcome assessment. Indeed, the overall performance score was quite high after reading both the AO ($M = 51.94$, $SD = 6.00$) and non-AO ($M = 50.93$, $SD = 6.32$). Upon further scrutiny, such high performance was likely caused by the fact that the multiple-choice items were too easy (median multiple-choice item sum score across both groups = 48, min = 31, max = 59, maximum possible score = 60). Such tendencies towards ceiling effects were not found for the short answer items (median multiple-choice item sum score across both groups = 4, min = 0, max = 8, maximum possible score = 8). To avoid ceiling effects, wherever possible, multiple-choice items (for example, "The set of conventions of a writing system are called…" with the answer alternatives "An orthography", "A phonics system", "A phonology" and "A morphology") were converted to short text answer format (for example, "Complete this sentence: The set of conventions of a writing system are called _____.").

Statistical assumption checks

Before conducting the ANCOVA, assumptions were checked. A Shapiro-Wilk test was used to determine whether residuals in the model were normally distributed. The results of the Shapiro-Wilk test indicated that the residuals did not significantly deviate from normality, $W = 0.99$, $p = .462$. This is further shown in the model diagnostic plots in Figure S4. The results of the Bartlett test indicated that the assumption of homogeneity of the residual variances was met, Bartlett's $K^2 = 0.07$, $p = .791$.

To determine linearity between the covariate and the outcome variable overall and at each level of the independent variable (that is, introductory material manipulation), scatter plots were created (see Figure S5). The scatter plots show the covariate (that is, standardised score in the prior knowledge assessment) on the x-axis and the dependent variable (that is, standardised score in the learning outcome assessment) on the y-axis. For the grouped scatter plot, data is shown in separate panels, depending on the kind of introductory material presented. Regression slopes predicting the dependent variable using the covariate were added to check the assumption of homogeneity of regression slopes. As shown in Figure S5, the regression lines are parallel. The lack of an interaction between the independent variable and covariate was further confirmed by a non-significant interaction term in the model, $F(1, 130) = 1.82$, $p = .180$ (see Table S3).

Supplementary materials

Wording of advance organiser (AO)

Imagine that you were asked to teach someone to drive. Teaching someone to drive is challenging for two key reasons:

1. Driving involves a complex, interacting array of knowledge and skills.
2. Driving is not a capability that comes naturally to most people.

A sensible question to ask is this: how do effective driving instructors handle both of these challenges?

Let's consider the first challenge. Because driving involves a complex, interacting array of knowledge and skills, driving instructors need to understand the different elements of driving. This means breaking down the act of driving into individual components (e.g., knowing how to use the brakes, knowing how to react to other vehicles on the road, etc.) so that they can be understood and explained. This knowledge also helps driving instructors to know which components to focus their teaching on at any one time.

Let's consider the second challenge. Because driving is not a capability that comes naturally to people, driving instructors need to ensure that learner drivers are given plenty of practice in every component of driving – and plenty of practice in integrating those components. Otherwise, the difficulties of learning to drive can be overwhelming.

What has this all got to do with teaching reading? Teaching someone to read has a lot in common with teaching someone to drive. Like driving, reading involves a complex, interacting array of knowledge and skills. Just like teaching someone to drive, this means that it is useful to break down the act of reading into individual components so that they can be understood and explained. Again, this knowledge helps teachers to know which components to focus on at any one time. Equally, reading is not a capability that comes naturally to most people. This means that teachers need to ensure that pupils are given plenty of practice in every component of reading – and plenty of practice in integrating those components. In short, reading instruction is most effective when teachers understand the different integrated components that underpin reading and when they ensure that pupils receive plenty of practice.

Wording of non-advance organiser (non-AO)

The development of literacy and the modern world as we know it have gone hand in hand. This can be seen as the result of two processes that interacted with each other:

1. The increasing complexity of societies made it necessary to develop writing systems and to educate increasing numbers of people who could use these writing systems.
2. The information stored and shared through these writing systems enhanced the increasing complexity of societies.

A sensible question to ask is this: how did these processes develop and how did they interact?

Let's consider the first process. Because societies increased in their complexity, the information that was represented using symbols became more complex. For instance, it was necessary to keep records of agricultural surplus and of trades. To represent this complex information, the writing systems had to become more complex. Simply representing objects with pictures was not enough. Over time, writing systems became more complex as they could portray every word in a spoken language. Writing systems were used for diverse purposes, making it necessary for them to be learned by increasing numbers of people. In the modern world as we know it, reading is seen as a valuable ability. Indeed, the last few centuries have seen a drive towards universal literacy in much of the world.

Let's consider the second process. Writing systems developed initially for a limited set of purposes. Yet, the new breadth and depth of information that they could convey changed the world. Complex information could be shared with following generations not merely through oral traditions, but through detailed writing systems. The ideas and works of scientists, artists and religious figures could be stored indefinitely. The invention of the printing press brought this information to ever larger numbers of people. This further enhanced the technological development of the modern world as we know it.

What has this all got to do with reading as we understand it? Reading is a foundational ability that allows people to engage with the modern world, not least due to the role of literacy in education.

Items included in learning outcome assessment

1. In terms of reading fluency, what word is used to describe the extent to which words are read aloud with appropriate stress and intonation? [short answer item; correct solution: **prosody**]

2. Which pupils might be best supported by prioritising the teaching of how and when to summarise what they read and how and when to ask questions about what they read? (Select one or more options.) [multiple-choice item; correct solution: **(c)**]

   (a) Pupils who lack the vocabulary to read academic texts.
   (b) Pupils who recognise individual words adequately, but who struggle to read fluently.
   (c) Pupils who appear to be relatively fluent readers but who pay little attention to building meaning as they read.
   (d) Pupils who struggle to decode words.

3. How many phonemes can be identified in the word 'splash'? [short answer item; correct solution: **5**]

4. A text says, 'Harriet kicked the ball into the net and ran off with her arms in the air.' A reader reasons that Harriet is likely to be playing football. This is an example of an inference because... (Select one or more options.) [multiple-choice item; correct solution: **(b)**]

   (a) It requires the reader to ignore some of the words and focus on others.
   (b) It requires the reader to draw a conclusion that isn't literally stated in the text.
   (c) It requires the reader to read sentences that came before this one.
   (d) It requires the reader to recognise every word.

5. A literacy lead and a science lead are adapting the science curriculum to support pupils' reading development. Which of these are likely to be elements of such a curriculum that support pupils' reading development? (Select one or more options.) [multiple-choice item; correct solution: **(a) (b) (c)**]

   (a) Carefully sequenced content to help build knowledge of the world.
   (b) Selected key scientific words to be taught to pupils.
   (c) Texts embedded into the curriculum that provide opportunities for reading practice.
   (d) Opportunities to draw inferences from pictures.

6. Complete this sentence: The smallest chunk of meaning within a word is called a _____ . [short answer item; correct solution: **morpheme**]

7. Complete this sentence: A morpheme placed on the end of a word to change it to a different word is called a _____ . [short answer item; correct solution: **suffix**]

8. Complete this sentence: Brief teaching of different _____ can be used to ensure that pupils are consciously aware that reading is an active constructive process. [short answer item; correct solution: **(comprehension) strategies / active comprehension**]

9. Complete this sentence: Frameworks related to reading tend to show the necessity of two components - word recognition and _____ . [short answer item; correct solution: **(language) comprehension**]

10. Complete this sentence: The approaches to reading and writing that are tailored to the purposes, methods and content of a particular area of academic study are called _____ literacy. [short answer item; correct solution: **disciplinary literacy / subject-specific literacy**]

11. How many graphemes are there in the word 'bracket'? [short answer item; correct solution: **6**]

12. Which of these are examples of comprehension strategies that might be used by pupils to support them to build meaning from a challenging text? (Select one or more options.) [multiple-choice item; correct solution: **(a) (b)**]

(a) Summarising.
(b) Clarifying.
(c) Blending.
(d) Highlighting.

13. Complete this sentence: The process of reasoning when reading that goes beyond the literal meaning of the text is called _____ . [short answer item; correct solution: **inference / inferencing / (verbal) reasoning**]

14. Complete this sentence: A pupil's comprehension of a text appears to be weak, even though their knowledge of vocabulary, knowledge of sentence structures, knowledge of text structure and background knowledge relevant to the text are all adequate, and they understand the need to read strategically and make inferences. The pupil also exhibits no particular weaknesses in their listening comprehension. The poor comprehension of text is likely to be the result of weak _____ . [short answer item; correct solution: **decoding / word recognition / fluency**]

15. Complete this sentence: When a phonics curriculum is organised so that it starts with the simplest, most common correspondences between phonemes and graphemes and incrementally adds complexity, this curriculum is said to be organised in a _____ fashion. [short answer item; correct solution: **systematic**]

16. As a result of extensive decoding practice, the quality of in-classroom discussions of new texts noticeably increases. How might this be explained? (Select one or more options.) [multiple-choice item; correct solution: **(b)**]

(a) Recognising the correspondence between graphemes and phonemes becomes less automated and error prone.
(b) Fluency has improved, meaning pupils have more working memory available to dedicate to building meaning from the texts.
(c) The pupils have learned that they do not need to accurately recognise the words in order to make meaning from the text.
(d) The pupils' level of prosody has reduced.

17. Complete this sentence: In English orthography, graphemes are used to represent the smallest recognisable units of sound that can be used to distinguish between two words. These units of sound are called _____ . [short answer item; correct solution: **phonemes**]

18. Complete this sentence: The set of conventions of a writing system are called _____ . [short answer item; correct solution: **(an) orthography**]

19. A pupil takes two reading comprehension assessments on the same day. The pupil achieves two very different scores on these assessments. This suggests that... (Select one or more options.) [multiple-choice item; correct solution: **(b) (d)**]

- **(a)** The pupil improved their reading fluency significantly.
- **(b)** The tests required different vocabulary and background knowledge.
- **(c)** Reading comprehension tests never provide useful information.
- **(d)** The results of individual comprehension assessments should not be treated as a definitive guide to a pupil's reading ability.

20. Which of these statements about the teaching of reading comprehension are correct? (Select one or more options.) [multiple-choice item; correct solution: **(d)**]

- **(a)** Comprehension of text is facilitated by reading fluency, but comprehension does not facilitate reading fluency.
- **(b)** Subject-specific approaches include exposure to a variety of text types and aim to increase the breadth, but not depth, of pupils' vocabulary.
- **(c)** The teaching of reading comprehension should occur in key stages 1 and 2, but not in key stages 3 and 4.
- **(d)** Reading comprehension lessons should not be entirely based upon answering questions in writing as this reduces the amount of time spent reading texts.

21. Complete this sentence: With extensive practice, the process of word recognition develops a sense of flow, which can be identified through the accuracy, automaticity and prosody of oral reading. This flow of words is called _____ . [short answer item; correct solution: **(reading) fluency**]

22. How many morphemes are there in the word 'regained'? [short answer item; correct solution: **3**]

23. Inferences are... (Select one or more options.) [multiple-choice item; correct solution: **(a) (b)**]

- **(a)** Central to the construction of meaning when reading texts.
- **(b)** Made by readers when they reach conclusions that go beyond the literal meaning of the text.
- **(c)** Only made by a person when they are reading.
- **(d)** Required only when reading narrative texts.

Items included in learning experience questionnaire

1. Overall, I was satisfied with the introductory material presented ahead of the evidence summary in module 1 of course 3 ('Learning to read').

   - ☐ Strongly disagree
   - ☐ Disagree
   - ☐ Somewhat disagree
   - ☐ Neutral
   - ☐ Somewhat agree
   - ☐ Agree
   - ☐ Strongly agree

2. Overall, I was satisfied with the evidence summary presented in module 1 of course 3 ('Learning to read').

   - ☐ Strongly disagree
   - ☐ Disagree
   - ☐ Somewhat disagree
   - ☐ Neutral
   - ☐ Somewhat agree
   - ☐ Agree
   - ☐ Strongly agree

3. How much mental effort did you invest to learn the content from the introductory material and evidence summary presented in module 1 of course 3 ('Learning to read')?

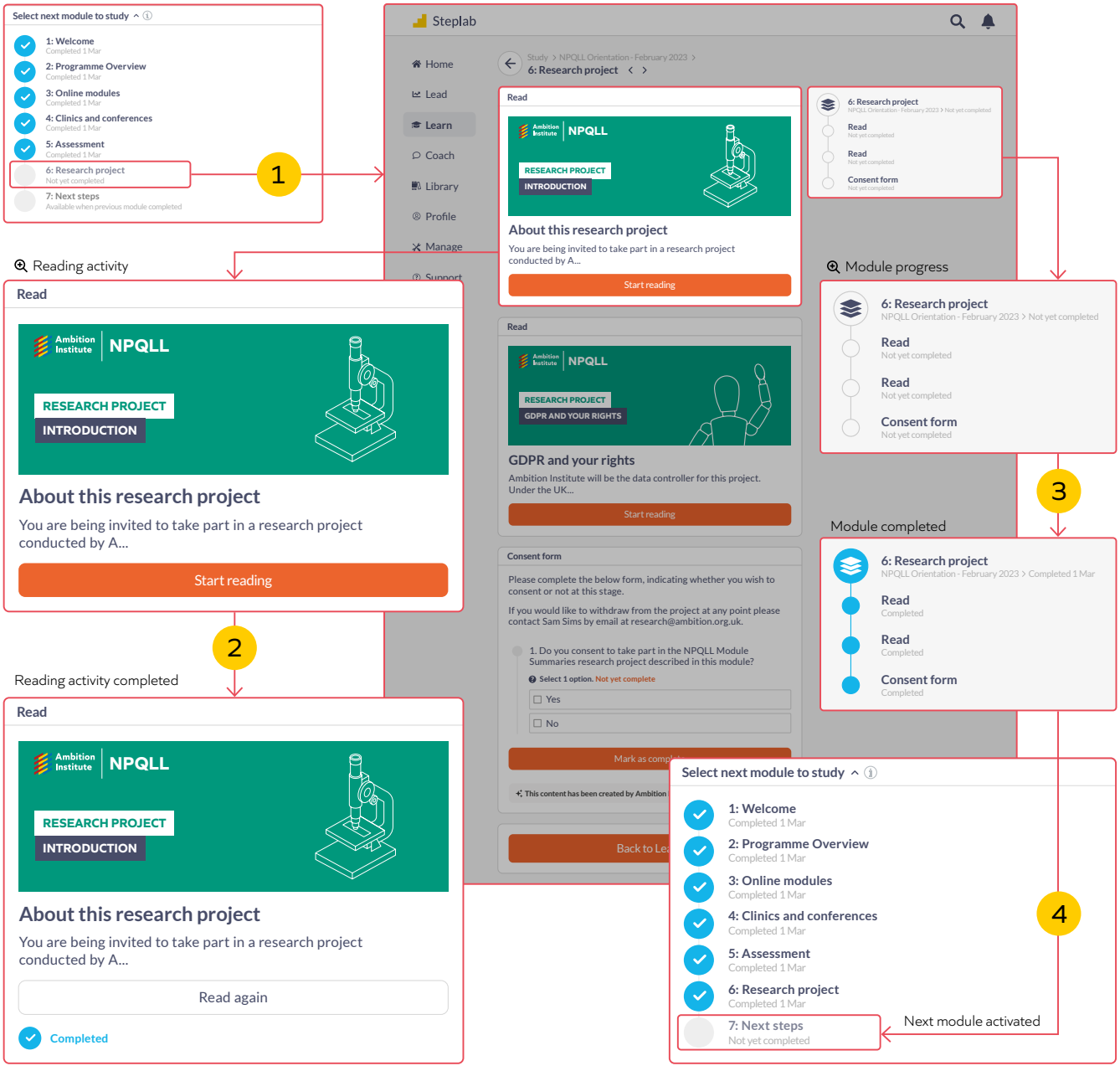
   - ☐ Very, very low mental effort
   - ☐ Very low mental effort
   - ☐ Low mental effort
   - ☐ Neither low nor high mental effort
   - ☐ High mental effort
   - ☐ Very high mental effort
   - ☐ Very, very high mental effort

4. How difficult was it for you to learn the content from the introductory material and evidence summary presented in module 1 of course 3 ('Learning to read')?

   - ☐ Very, very easy
   - ☐ Very easy
   - ☐ Easy
   - ☐ Neutral
   - ☐ Difficult
   - ☐ Very difficult
   - ☐ Very, very difficult

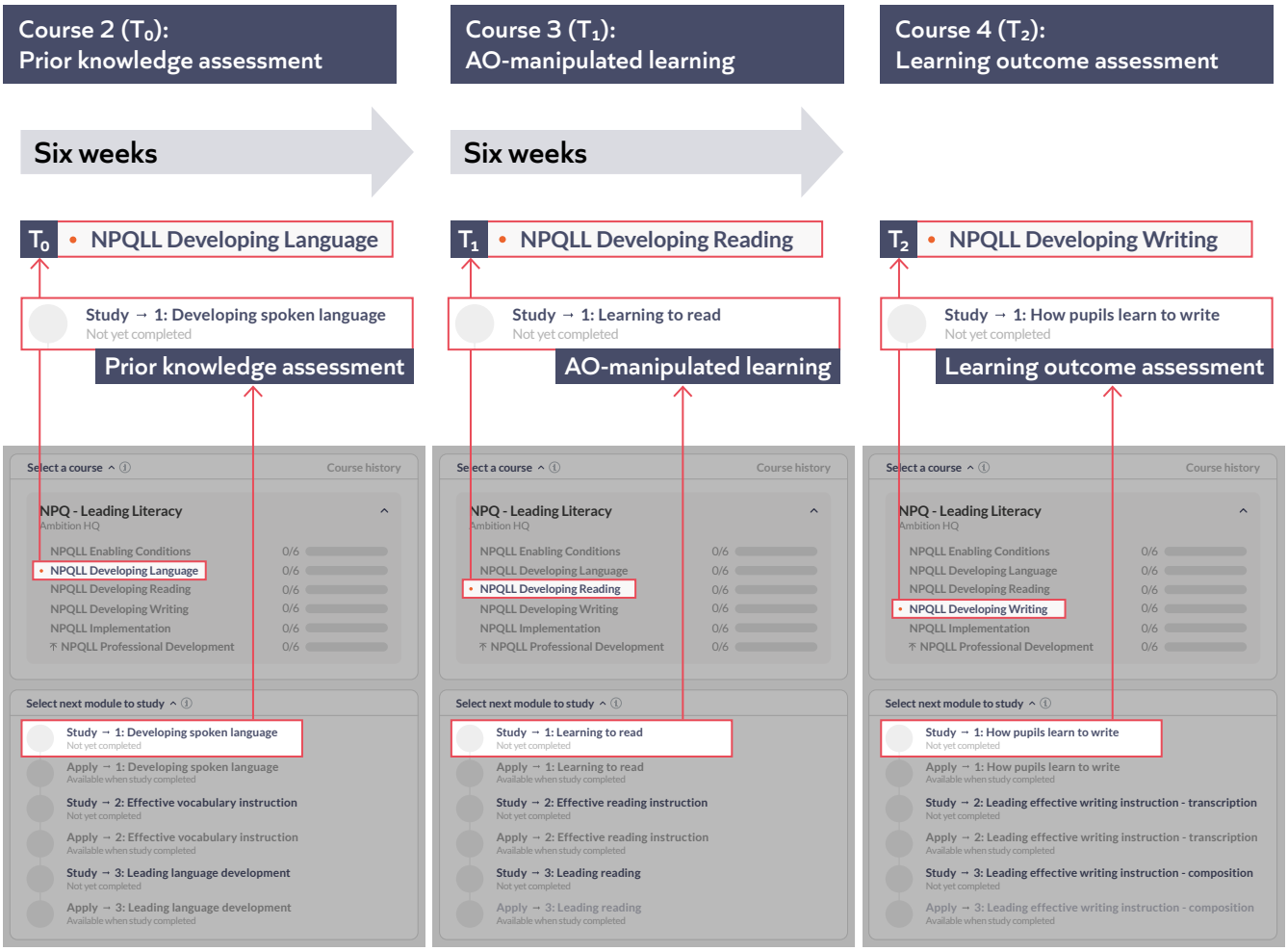
Supplementary figures

Figure S1. Participant recruitment and consent embedded in the orientation course



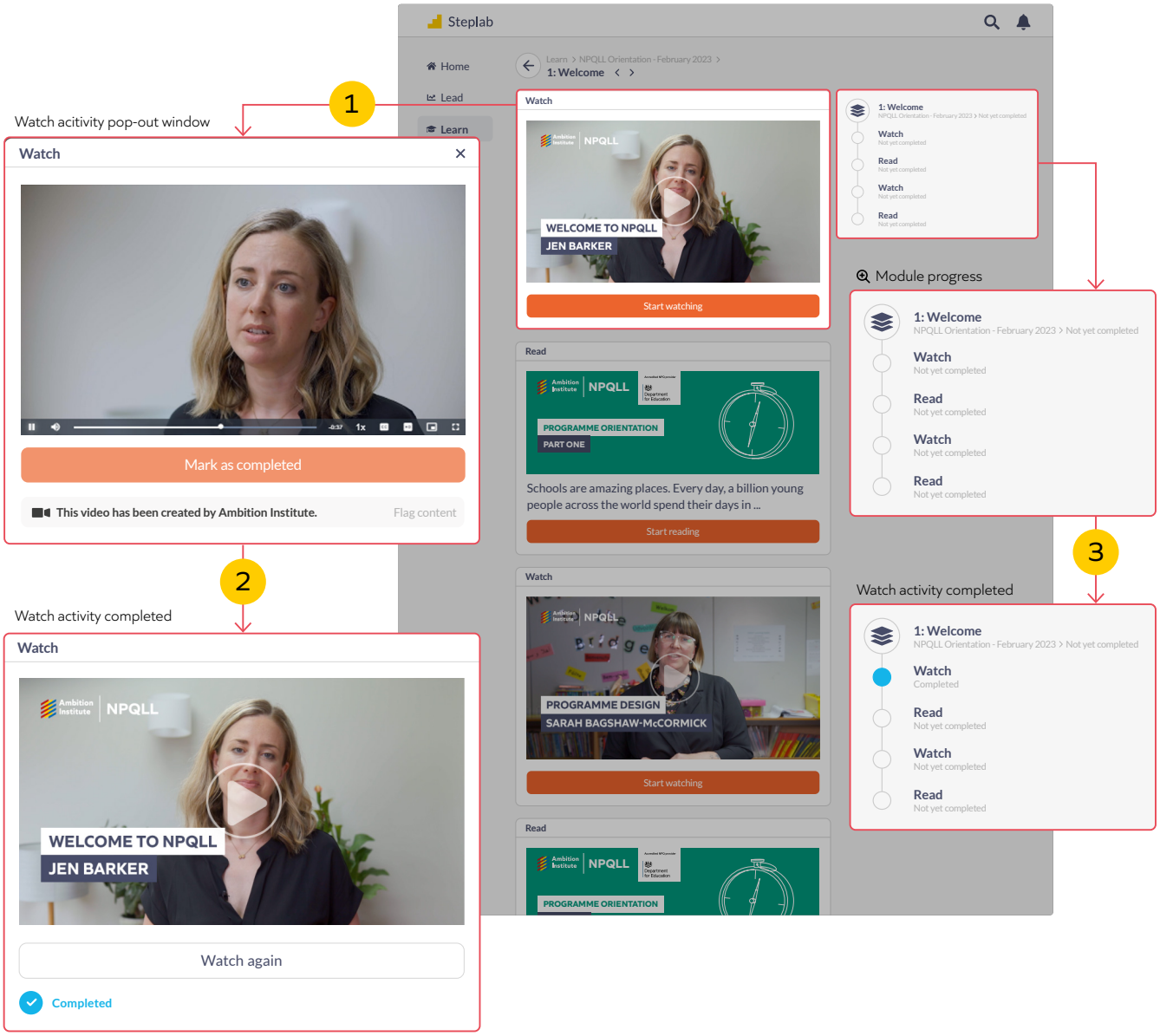
Note. The figure includes screenshots from the Steplab platform. The orientation course consisted of seven modules that must be completed in consecutive order. The completion of a previous module grants access to the next module. Information about the research project was added as an independent module and consisted of the information about the study and General Data Protection Regulation, as well as a consent form.

Figure S2. Content of courses and modules used in the research project



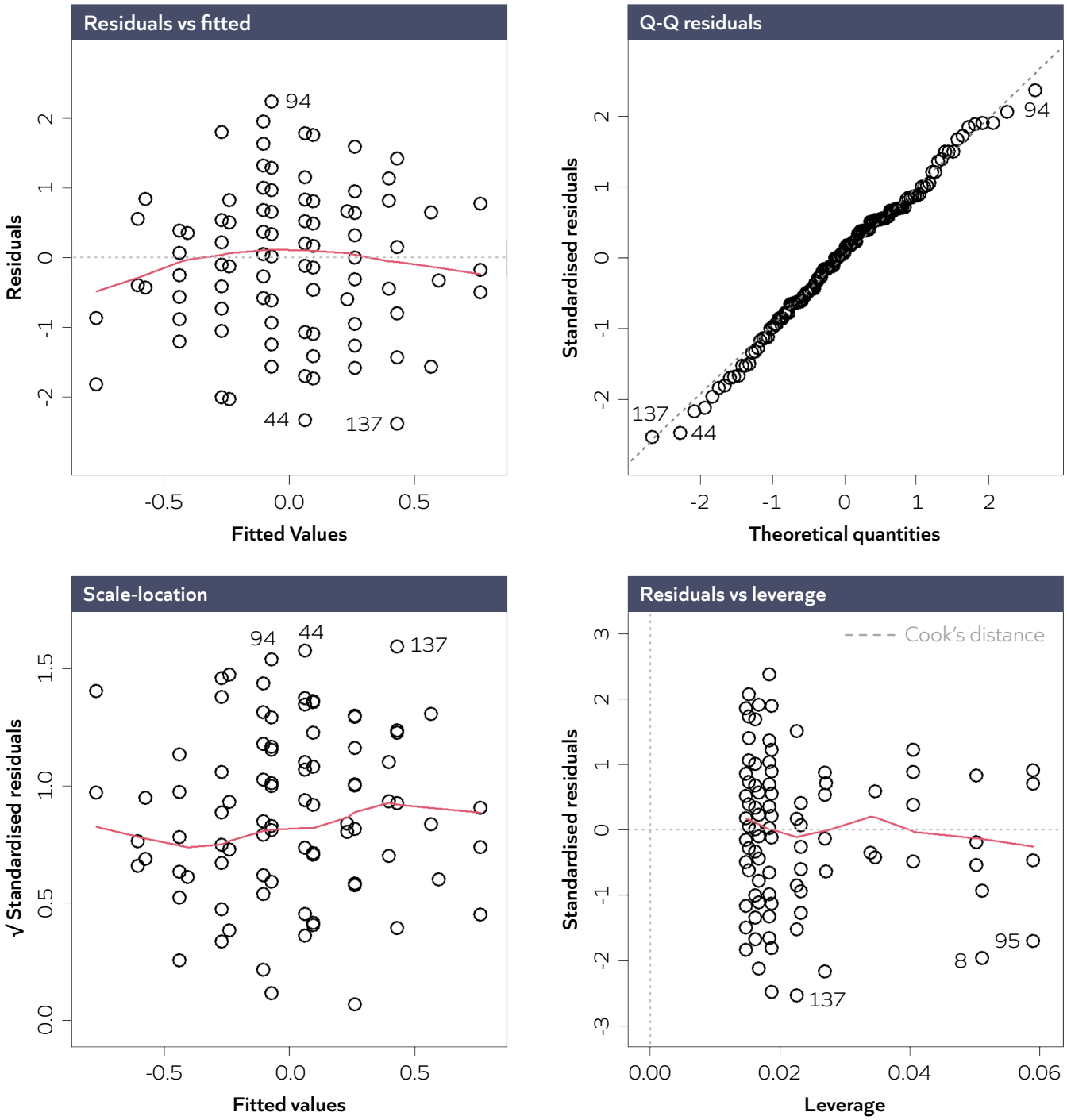
Note. This figure corresponds to Figure 2 in the main text. The three online courses modified for research purposes are shown. The courses were released sequentially every six weeks. Each course consisted of three module pairs that are listed below. "Apply" modules only became available after the corresponding "Study" module had been completed. Research objects were embedded in the first "Study" module of each course (highlighted in coral).

Figure S3. Objects embedded in a module



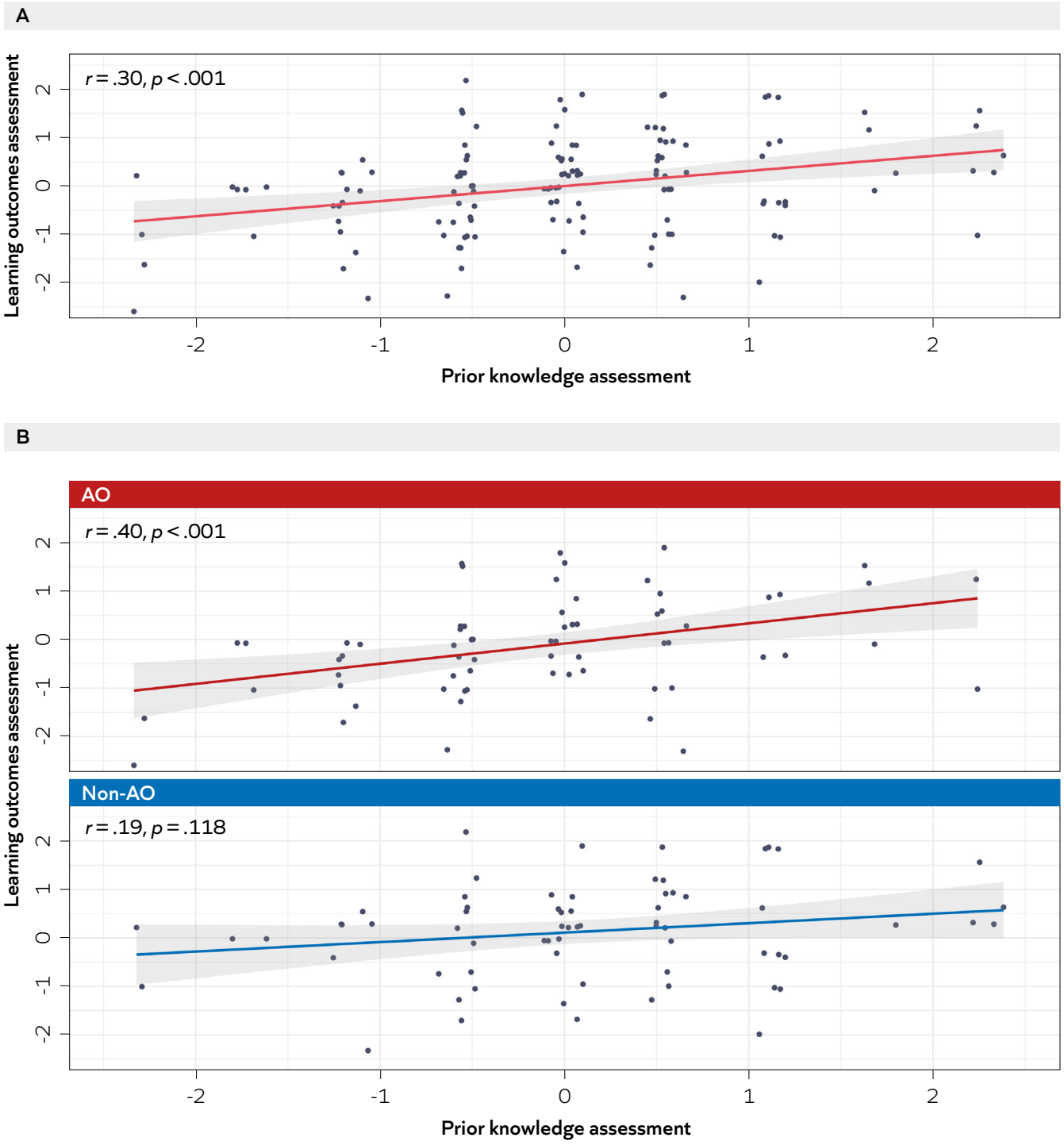
Note. The figure shows screenshots of the Steplab platform. As an example, we here used the "NPQLL Orientation" Course and the "Welcome" Module therein. A module contains multiple objects, here labelled as "Watch" or "Read". The objects are also listed in the progress view on the right-hand side of the screen with corresponding labels (see dashed arrows). The screenshots and arrows illustrate how participants can engage with an object, here we show an example of a video object. Upon clicking "Start watching", the object is opened and appears highlighted in the centre of the screen. This also collects a timestamp of when the object was accessed. When participants finish engaging with the content of the object, the object can be "Mark[ed] as completed". This then closes the highlighted object view and the "Start watching" field is replaced with "Watch again" together with a stamp showing that the object is completed below the object and in the progress view (in cyan). Additionally, marking the object as completed overwrites the initial timestamp of when the object was accessed with a timestamp of when it was completed.

Figure S4. Diagnostic plots



Note. The figure contains the plot diagnostics for the fitted linear model as recreated with the '*plot.lm()*' functionality in R. The top-left plot ('Residual vs fitted' plot) shows if the residuals have non-linear patterns. The top right plot ('Residual Q-Q' plot) shows if the residuals are normally distributed. The bottom left plot ('Scale-location' plot) shows whether the variances of the residuals are homoscedastic. The bottom right plot ('Residuals vs leverage' plot) shows any influential cases in the linear model.

Figure S5. Scatterplot of the prior knowledge covariate and dependent variable



Note. In the scatter plots, the covariate prior knowledge assessment is plotted on the x-axis and the dependent variable learning outcome assessment is plotted on the y-axis. Both variables were standardised and the points were jittered to avoid overplotting. Regression slopes and 95% confidence interval are displayed as coloured lines and shaded areas. (A) shows the data of the whole sample whereas (B) shows the data for each group separately.

Supplementary tables

Table S1. Comparison of (psycho-)linguistic characteristics of the introductory materials

	AO	Non-AO
Readability consensus		
Grade level	10	10
Reading level	Fairly difficult to read	Fairly difficult to read
Reader's age in years	14-15 (Ninth to Tenth graders)	14-15 (Ninth to Tenth graders)
Readability consensus		
Total # of words	345	344
Total # of unique words	140 (41%)	171 (50%)
Total # of repeat words	205 (59%)	173 (50%)
Average # of words per sentence	14	14
Total # of sentences	24	24
Total # of characters	1749	1750
Total syllables in text	574	587
Total # of words with single syllables	210 (61%)	213 (62%)
Total # of words with double syllables	84 (24%)	75 (22%)
Total # of words with 3+ syllables	51 (15%)	56 (16%)
English Lexicon Project [*M* (*SD*)]		
Number of characters	5.14 (2.77)	5.09 (2.69)
Number of syllables	1.62 (0.85)	1.66 (0.98)
LgSUBTLEX frequency/1000000 words	4.33 (1.4)	4.3 (1.43)
RT lexical decision task (*z* score)	-0.55 (0.21)	-0.55 (0.21)
RT naming task (*z* score)	-0.49 (0.23)	-0.49 (0.23)
Age of acquisition	5.64 (2.11)	5.83 (2.31)
Concreteness rating	2.33 (0.91)	2.28 (0.86)
Emotional valence	6.1 (0.74)	5.98 (0.76)
Emotional arousal	3.78 (0.72)	3.91 (0.78)
Emotional dominance	5.94 (0.78)	5.73 (0.62)

Note. The statistics for readability consensus and word characteristics were generated using an automated readability checker, available at **readabilityformulas.com/free-readability-formula-tests.php**. The English Lexicon Project database (Balota et al., 2007) is available at **elexicon.wustl.edu/index.html**. In addition to standardised behavioural and descriptive data, available ratings of concreteness (Brysbaert et al., 2014) as well as emotional valence, arousal, and dominance (Warriner et al., 2013) were extracted for each word to calculate *M* (*SD*) shown here. # = count; LgSUBTLEX = $\log_{10}$ (number of times the word appears in the subtitle frequency corpus + 1); RT = reaction time.

Table S2. Items used in the prior knowledge assessment

Alternative 1	Alternative 2	Alternative 3	Alternative 4
Item 1. In terms of reading, inferences are conclusions that…			
Go beyond the literal meaning of the text	Don't require background knowledge	Necessarily require conscious thinking	Are not necessary when reading non-fiction texts
Item 2. Which of these comprehension strategies has the least sound basis in evidence?			
Summarising	Clarifying	Self-questioning	**Extrapolating**
Item 3. Alongside chunks of spoken sound, the English writing system represents the smallest chunks within words that are associated with meaning called…			
Morphemes	Lexemes	Syntactic identifiers	Graphemes
Item 4. A grapheme is an individual letter or a group of letters that represents…			
A syllable	**A phoneme**	The sound of a whole word	A morpheme
Item 5. Reading fluency is commonly identified through three properties of oral reading called…			
Accuracy, automaticity and semantics	**Accuracy, automaticity and prosody**	Accuracy, automaticity and syntax	Accuracy, automaticity and flow
Item 6. Which of these capabilities is least likely to contribute to a person's reading fluency?			
Word recognition	Syntactic knowledge	Vocabulary knowledge	**Knowledge of comprehension strategies**
Item 7. A phoneme can be thought of as…			
The chunk of sound associated with a morpheme	The chunk of sound associated with a syllable	**The smallest chunk of spoken sound that can distinguish one word from another**	Any chunk of sound that is smaller than a whole word
Item 8. An orthography is…			
The set of conventions associated with a phonics programme	**The set of conventions associated with a written language**	The set of conventions associated with a handwriting system	The set of conventions associated with a spoken language
Item 9. Reading comprehension can be described as relying on two capacities: the ability to recognise words and build meaning from words. These are generally described as…			
Word recognition and background knowledge	**Word recognition and language comprehension**	Word recognition and reading comprehension	Word recognition and fluency
Item 10. Vocabulary tiers form a construct that is commonly used to prioritise particular words to teach. Tier 3 vocabulary can be thought of as…			
Basic words that pupils are likely to learn through their everyday interactions	Words that are high frequency for mature language users and are found across a variety of domains.	Words that are found primarily in scientific subjects	**Words that are quite low frequency and restricted to specific domains**

Note. Items 4, 7, 8, 9, and 10 were modified following the pilot study. The correct answer for each item is highlighted in bold.

Table S3. ANCOVA results with interaction term

	SS	*df*	MS	*F* value	*p* value	η_p^2
AO effect	1.3	1	1.3	1.44	.233	0.01
Prior knowledge	11.24	1	11.24	12.41	.001	0.09
AO x prior knowledge	1.65	1	1.65	1.82	.180	0.01
Residuals	117.71	130	0.91			

Note. SS = sum of squares, *df* = degrees of freedom, MS = mean square, η_p^2 = partial Eta squared.

A great teacher changes the future every day. They can be the critical factor in a child's success, especially for those who have had a tough start in life. At Ambition Institute, we support teachers and school leaders at every stage of their careers, helping them to keep getting better. We are a charity providing training and professional development based on the most rigorous research and evidence about what really works. Together, we're shaping the future of education to give every child the best start in life.

Learn more

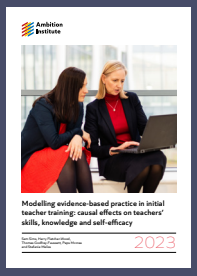
Get in touch today to find out how you can nominate your staff, or yourself, for one of our programmes.

ambition.org.uk
info@ambition.org.uk
020 3668 6865
@Ambition_Inst

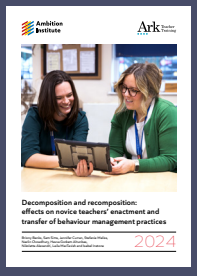
Ambition Institute
2 Bridge Wharf,
156 Caledonian Road,
London, N1 9UU

Discover more

Evidence is at the heart of what we do at Ambition Institute. Through our research we want to push forward our collective knowledge of effective professional development and share what works, to turn evidence into action. We aim to answer important questions faced by the sector, help teachers and leaders to address disadvantage, and show the impact of evidence-informed teaching.



Does adding modelling to professional development help to bridge the theory-practice gap? Our study looked at the practice of initial teacher trainees who used video models to develop their retrieval practice skills.



This research explores whether breaking down (decomposing) and then recombining teaching practices into a different sequence of teaching (recomposition) can help teachers apply what they have learnt to new contexts.

Ambition Institute

© Ambition Institute 2024. Ambition Institute is a registered charity, number 1146924.