

The impact of multi-word units in early foreign language learning and teaching contexts: A systematic review

Johannes Schulz  | Catherine Hamilton | Elizabeth Wonnacott | Victoria Murphy

Department of Education, University of Oxford, Oxford, UK

Correspondence

Johannes Schulz, Department of Education, University of Oxford, 15 Norham Gardens, OX26PY Oxford, UK.

Email: johannes.schulz@education.ox.ac.uk

Abstract

This systematic review reports on research investigating the impact of multi-word unit (MWU) input on young learners' second language (L2) attainment in instructed settings. Recent findings suggest that L2 learners can generalise from MWU input, abstract patterns and employ such schemata productively via slot-filling, indicating that MWUs are key catalysts of learners' L2 development. Simultaneously, primary school L2 instruction is on the rise worldwide and the importance of MWUs is acknowledged in curricula, teacher education and teaching materials. Therefore, the incentive of this review is to systematically report the state of the art of research regarding the impact of MWU instruction in early L2 teaching contexts. The review covers English, German and French research into typically developing monolingual children aged 5–12 learning an L2 in instructed teaching settings. Only two of the total results ($n=2233$) met the inclusion criteria. Following quality assessment using the Mixed Methods Appraisal Tool and based on a narrative synthesis of available results, we cannot report trustworthy evidence of the effectiveness of teaching MWUs to young L2 learners. We highlight the lack of research evidence and conclude that existing research lacks robust evidence that MWU input already established in teaching contexts has a measurable effect on specific aspects of students' L2 attainment, such as productive skills. Although we promote

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Review of Education* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

MWU's potentially facilitating role in L2 development, we call for more classroom-based intervention research on MWUs in primary school contexts to enable much-needed evidence-based recommendations for L2 teaching to support L2 learning outcomes in primary schools.

KEYWORDS

foreign language, language acquisition, multi-word units, primary school, systematic review

Context and implications

Rationale for this study

Research has shown that multi-word units (MWUs) are key catalysts for L1 and L2 learning. MWUs are crucial to current primary school L2 teaching and their importance is acknowledged in curricula. This review systematically examines the scientific evidence for the effectiveness of MWU instruction on L2 attainment.

Why the findings matter

Our review cannot report trustworthy evidence of the effectiveness of teaching MWUs to young L2 learners. This lack of evidence is striking since early L2 instruction is on the rise worldwide and MWUs are already integral parts of L2 curricula, teacher education and teaching materials.

Implications for the research community and policy makers

Given our findings, we call for more research in instructed settings to better understand the impact of MWU input on specific aspects of students' L2 attainment, such as productive skills. Based on such work, researchers could provide policy makers and practitioners with evidence-based recommendations that help integrate MWU input into L2 curricula and teaching in such a way that is most beneficial for young students' L2 attainment.

INTRODUCTION

This systematic review reports on research investigating the effectiveness of multi-word unit (MWU) input in instructed foreign language (FL) settings with young learners. Research suggesting that MWUs are integral building blocks of language learning and processing for first language (L1) acquisition (Bannard & Lieven, 2012; Theakston & Lieven, 2017) and second language (L2) learning (Eskildsen, 2009; Myles et al., 1999) has introduced many conceptualisations of MWUs, including *prefabricated routines*, *patterns*, *formulaic sequences*, *chunks* and *multi-word units* (Christiansen & Arnon, 2017; Czarnecka, 2011; Wray, 2002, 2013). As this review focuses on MWU's facilitating role for language learners, we consider that MWUs 'provide children with lexically specific chunks to be used in early production and

allow them to discover distributional and structural relations that hold between words' (Arnon & Christiansen, 2017: 622).

Since the 1980s, English as a Foreign Language (EFL) classes at primary level have been introduced worldwide (Enever, 2011; Graddol, 2006; Rixon, 2013). Europe joined this movement in the early 2000s (Council of Europe, 2001), and the UK made L2 classes mandatory from KS2 in 2014 (Holmes & Myles, 2019). In primary school contexts, current L2 curricula focus on communicative interactions and metalinguistic competences (Kultusministerium Baden-Württemberg [KM-BW], 2004: 68–71), aiming to make children use the L2 productively and to educate active language users (Bredenbröcker, 2018; Diehr, 2009; Sambanis, 2007). Precisely to provide children with generative patterns that allow rule extraction and equip them with early communicative agency, MWUs have been established in European L2 curricula (Council of Europe, 2018; Europarat, 2001), as in Germany for example (e.g., *chunks*, *formulaic sequences*; see KM-BW, 2016). Correspondingly, MWUs are part of some official teaching materials (e.g., *formulaic speech*, *chunks* in Große-Brauckmann & Heidelberg, 2018) and teacher education programmes (Legutke et al., 2012).

Despite the rise of L2 instruction in primary schools and the prevailing role of MWUs in those contexts, selective literature reviews we conducted for teaching intervention studies with a focus on the *effectiveness* of targeted MWU input in primary school L2 settings yielded few results. There appears to be a shortage of studies researching the effectiveness of such input on relevant linguistic outcome variables, such as slot-filling abilities. However, we argue that MWUs rightly have made their way into curricula and classrooms, and that the role of MWUs in these contexts needs to be researched. In the following, we explore why MWUs are promising for instructed settings by discussing theoretical underpinnings and research evidence supporting the facilitating role of MWUs in naturalistic and instructed learning settings.

Previous research

From a theoretical perspective, MWU's role is accounted for in constructivist frameworks such as Usage-Based Language (UBL) learning theory, which maintains that learners induce communicative functions and structural information of linguistic constructions from input embedded in specific communicative situations (Behrens, 2009; Tomasello, 2003). Across multiple encounters, construction-specific form-function mappings develop; learning is therefore considered exemplar-based (Ellis, 2002). Via gradual pattern abstractions across usage-events, learners' linguistic inventory, additionally impacted by input characteristics including salience and frequency, emerges along a continuum of construction-abstractness. This continuum varies from concrete units [*I do not know*], via semi-fixed formulas [*I do not + verb*], to abstract constructions like syntactic schemas [*Subj Verb (Obj)*] (Ibbotson, 2013).

In L1 research, so-called 'traceback' studies—where researchers track MWUs in child corpora—support this incremental linguistic inventory development (e.g., Lieven et al., 2003, 2009; Peters, 1977). For example, Lieven et al. (2009) report that 2-year-old children's MWUs gradually become more schematic with more complex slots. Moreover, psycholinguistic work demonstrating that adults process early acquired MWUs faster than later acquired ones (Arnon et al., 2017) and that 1-year-olds are sensitive to the frequency of *both* single words (Bergelson & Aslin, 2017) and MWUs (Skarabela et al., 2021) supports the important early role of MWUs in language acquisition (Bannard & Lieven, 2012; Theakston & Lieven, 2017).

Regarding L2 learning, as early as 1938, Kenyeres & Kenyeres (1938) tracked their Hungarian-speaking daughter's untutored L2 French learning in Geneva, finding that she

frequently imitated memorised MWUs to communicate. Following an argument with her mother after 2 months of exposure, the child produced the following recombination of individual MWUs: Maman, s'il vous plaît, qu'est-ce que c'est, voulez-vous? (Mummy, please, what is it, would you like [to])? After 3 months, she began to modify elements in MWUs relative to the communicative context, such as changing third person singular verbs to plural. Decades later, Fillmore (1976) tracked five Mexican children's L2 English development, reporting that they gradually extracted individual constituents from MWUs according to communicative context. Once all constituents of a MWU had been extracted, the children were left with an abstract grammatical structure that they could use generatively. Other traceback studies confirm that children's naturalistic L2 acquisition heavily relies on MWUs (Hakuta, 1974; Huang & Hatch, 1978). These early studies demonstrate key characteristics of MWUs' role in L2 learning: (a) MWUs equip learners with early communicative agency and (b) constitute developmental 'starting points' for generalisation and abstraction processes.

Building on naturalistic L2 development findings, research into instructed L2 learning has become prominent in recent years (Ellis, 2015; Ellis & Ferreira-Junior, 2009; Ellis & Wulff, 2020; Eskildsen, 2009; Eskildsen & Cadierno, 2015; Mellow, 2006; Roehr-Brackin, 2014; Tyler, 2010). For example, Eskildsen (2009) traced the use of *can*-patterns in longitudinal data collected from one adult Spaniard (Carlos) learning English as a Second Language (ESL). The data included 120 h of aural recordings of Carlos in ESL classrooms over 4 years. Counting MWU occurrences in the participant's production data, Eskildsen distinguished a 'starting' pattern (*I can* + verb) based on which Carlos gradually developed more sophisticated constructions, such as *can you/can I/you can* + verb. These findings correspond to work in naturalistic L2 settings, suggesting that new and more abstract patterns emerged from concrete units.

In classrooms with younger learners, Myles et al. (1999) investigated MWU's role in students' creative construction of L2 interrogatives. Over 2 years, the authors collected data from 60 beginning French L2 learners in Year 7 (age 11) in England. Like Eskildsen (2009), Myles et al.'s data confirm that interrogative chunks such as *comment t'appelles-tu* (*what's your name?*) were catalysts for linguistic analysis and free production. For example, late in the study, one student used structural information from an initial chunk (*comment t'appelles-tu?*) as a schematic facilitator for advanced utterances like *comment s'appelle le garçon/fille* (*boy/girl?*)? Those findings suggest that young learners, too, can learn an L2 from initially unanalysed chunk input that they gradually segment and analyse. Over time, this process leads to increasingly productive use of abstract structural knowledge (Kersten, 2015).

Since research suggests that MWUs are beneficial to L2 learners in naturalistic and instructed settings, it is logical to implement MWUs into L2 classroom teaching to facilitate learning, particularly to investigate whether and to what extent implementation will improve learning. In fact, research has investigated MWUs in instructed non-beginner adult L2 contexts (Boers et al., 2006, 2017; Boers & Muñoz-Basols, 2021; Choi, 2017; Eyckmans & Lindstromberg, 2017; Thomson, 2020; Wood, 2009; comprehensive overview: Boers & Lindstromberg, 2012). However, such work has mainly focused on fluency (Thomson, 2020), idiomaticity (Boers & Muñoz-Basols, 2021) and processing speed (Choi, 2017). Here MWUs are not considered developmental 'starting points' because participants are often proficient, or at least past the initial learning stages. This emphasises an important distinction between MWUs as drivers of fluency and idiomaticity and as drivers of language acquisition (Aguado, 2002). Since MWUs are sources for grammatical knowledge extraction, they can be catalysts of language learning at the *beginning* of learning trajectories (Kersten, 2015).

It follows then that *young* L2 learners especially might benefit from targeted MWU input to launch their L2 development. Primary school L2 instruction has been introduced worldwide and some contexts have implemented findings supporting MWU's facilitating role in language learning into primary L2 curricula and classrooms (see [Introduction](#)). Interestingly,

results from some evaluations of primary school L2 learning outcomes reflect MWU's important role in students' L2 learning. A relevant example of such research is found within the context of Germany. The following section, therefore, discusses evaluations in Germany, highlighting the role of MWUs in young learners' L2 development.

Several studies in German primary schools report on children's L2 attainment. Engel et al.'s (2009) large-scale assessment of L2 attainment at the end of Year 4 (age 9–10) found that 42.2% of learners mainly reproduced MWUs in their communication, while 24.3% could sometimes vary individual units of MWUs. The so-called BIG-study (BIG-Kreis, 2015; Müller et al., 2016), another large-scale assessment at the end of Year 4, reports that 79.5% of utterances in student dialogues consisted of holistically learned MWUs. Other studies focused on learners' learning *processes* rather than assessing learning *outcomes*. Werlen (2007) tracked the L2 progress of 12 primary school classes over 4 years (Years 1–4; beginner learners aged 6–10) through video-recordings of English/French classes, audio data transcriptions and interviews with individual learners. She reports that learners gradually extracted MWUs from the input and modified them for personal communication goals. Sambanis (2007) reports similar longitudinal action research data from Year 1 and 2 English and French classes, maintaining that some learners could vary MWUs via slot-filling. Kahl and Knebler (1996) report that MWUs provided weak students in Years 3 and 4 with linguistic frames to rely on while strong students used them to advance to independent productions. Diehr (2009) analysed speech data of 216 children from 10 classes in Years 3 and 4. She reports that by the end of Year 4, many students managed to variably fill slots in patterns, showing signs of a developing awareness for grammatical structure.

The current study

Evaluation findings such as those presented in the previous section highlight the potential of MWUs for achieving early communicative agency. They also conform to constructivist notions in previously discussed L2 studies, suggesting that MWUs constitute integral building blocks and catalysts for primary students' L2 development. However, discerning cause-effect relationships from teaching input to findings of evaluations is difficult in several respects. Firstly, studies such as Fillmore (1976) are situated in untutored immersion contexts, whereas evaluations report on instructed settings. Critically, learning context is important since L2 input in teaching settings cannot compare to the quantity and quality of input in immersion settings. Secondly, even in contexts where MWUs are established in L2 curricula, it is the individual teacher's decision how to implement them. Although recommendations about MWU teaching exist, such as advocating teaching useful high-frequency MWUs (Bredenbröcker, 2018), we typically do not know which MWUs were taught and how. This lack of detail makes it impossible to link evaluation findings to confident conclusions that MWUs are crucial in primary school L2 development. As discussed in the [Introduction](#), selective literature reviews for teaching intervention studies focusing on the effectiveness of targeted MWU input in primary school L2 settings were unfruitful. While many findings from L2 traceback, psycholinguistic and evaluation studies suggest that MWUs can facilitate early L2 learning, we lack robust classroom-based evidence that MWU input, already established in curricula, has a measurable effect on specific aspects of students' L2 attainment, such as productive skills. Without that evidence, it is not possible to offer evidence-based recommendations for L2 teaching.

We therefore conducted a systematic review to provide a rigorous overview of the work that has been done in this research area and to critically evaluate and synthesise relevant results. Understanding the effects of purposefully focusing on MWU input in early L2 teaching could be powerful for L2 researchers, stakeholders, and practitioners. Firstly, it

would provide researchers with evidence to better understand and explain children's L2 development findings from evaluation studies, which in many respects tally with UBL theory and naturalistic L2 learning trajectories. Secondly, it would help policy makers, teacher educators and teaching material developers to justify and potentially readjust the inclusion of MWUs in curricula, teacher education and materials. Finally, it would support teachers in knowing to what extent certain types of focused MWU input impact learning outcomes and how those might be improved. However, without the requisite research, the potential impact of teaching MWUs to young L2 learners in formal educational contexts remains guesswork. Consequently, this review addresses the following research questions:

RQ 1.	What is the extent of original research investigating the role of MWU input in early FL learning and teaching contexts?
RQ 2.	What is the extent of the evidence on the effectiveness of learning from/teaching MWUs in early FL contexts?
2a.	What impact does MWU input have on young learners' grammatical and vocabulary skills and knowledge in the target language?
2b.	What impact does MWU input have on young learners' communication abilities (e.g., quantity and quality of spoken output; receptive/productive vocabulary knowledge)?

METHODOLOGY

Eligibility criteria

This review was pre-registered on IDESR (International Database of Education Systematic Reviews; <https://idesr.org/article/IDESR000016>) and covered literature on teaching interventions with typically developing monolingual children aged 5–12 learning an L2 in instructed settings. Detailed inclusion and exclusion criteria are provided in Table 1. After unfruitful selective literature searches, we included all languages of publication and grey literature.

Information sources

The consulted databases covered education, linguistics, psychology and multidisciplinary sources. Since between us we speak German and French in addition to English, we increased search scopes by adding relevant German and French databases. All databases are shown in Table 2. Individual German and French databases are discussed below.

German

After consultation with German researchers, we conducted pilot searches on the *Fachportal Pädagogik*, *PSYINDEX*, and the Humboldt University zu Berlin's online catalogue. The latter was consulted because, unlike other university databases, it processes Boolean strings. Following pilot searches, we decided on the Humboldt University's catalogue and the *Fachportal Pädagogik*, run by the Leibniz Institute for Research and Information in Education. Being Germany's main Education Index, it processes Boolean strings and runs meta-searches in several databases such as *FIS Bildung Literaturdatenbank* and *BASE*. *PSYINDEX* was excluded because pilot searches mainly yielded results from psychology, such as psychometric tests.

TABLE 1 Eligibility criteria.

Item	Inclusion criteria	Rationale
Bibliographic information	<i>INCLUDE:</i> Studies with a full reference or sufficient bibliographic information <i>EXCLUDE:</i> Studies with insufficient bibliographic information	Without sufficient bibliographic information, retrieval of works is unfeasible
Date of publication	<i>INCLUDE:</i> all <i>EXCLUDE:</i> none	Preliminary scoping searches showed that research on this topic is scarce, therefore, all available work will be included
Population	<i>INCLUDE A:</i> Children aged 5–12 <i>EXCLUDE A:</i> Children under the age of 5 and over the age of 12 <i>INCLUDE B:</i> Learners of any foreign language (including but not limited to English as a Foreign Language) and artificial language, if evidence is available <i>INCLUDE C:</i> All target languages where the language is a taught, foreign language <i>EXCLUDE C:</i> Minority language learning contexts <i>INCLUDE D:</i> Studies on typically developing foreign language learners. Include studies even if no explicit reference is made to learning ability if reasonable assumption can be made that participants are comprised mainly of typically developing individuals <i>EXCLUDE D:</i> Studies that exclusively target non-typically developing learners	This research is concerned with young learners who are at primary school age. This is the age range where European foreign language curricula have targeted MWU input since the early 2000s This research is concerned with general, non-language-specific foreign language acquisition Minority language learners are exposed to the target language outside of learning and teaching settings (e.g., in their homes or home communities). This research, however, is concerned with 'foreign' language learners who learn a new language outside the target language community, in taught, input-limited contexts This review seeks to assess the effectiveness of providing early foreign language learners with MWU input in teaching and learning settings as applies to typically developing populations. The findings for non-typically developing individuals may not hold for a larger population, and so these results should not be extrapolated, nor will they be included in this review

(Continues)

TABLE 1 (Continued)

Item	Inclusion criteria	Rationale
Intervention	<p><i>INCLUDE A:</i> Studies involving interventions in teaching settings (such as classrooms or language clubs) with a focus on MWU input, addressing any (or all) of the four skills (i.e., reading, listening, writing, speaking), and studies involving correlational design with a focus on MWU input, e.g., studies investigating the correlation between reading comprehension and MWU input</p> <p><i>EXCLUDE A:</i> Studies where no particular emphasis is put on MWU in the input and where no particular emphasis is put on correlations between the phenomenon of MWUs in the learning trajectory and other linguistic variables</p> <p><i>INCLUDE B:</i> Studies involving interventional and correlational designs in experimental laboratory settings (e.g., priming or eye-tracking experiments) with a focus on MWU input, addressing any (or all) of the four skills (i.e., reading, listening, writing, speaking)</p> <p><i>EXCLUDE B:</i> Studies where no particular emphasis is put on MWU in the input and where no particular emphasis is put on correlations between the phenomenon of MWUs in the learning trajectory and other linguistic variables</p> <p><i>INCLUDE C:</i> Studies that define 'linguistic units larger than one word' with terms other than 'MWU', including but not limited to 'formulaic sequences', 'formulas', 'multi-unit expressions', or 'lexicalised phrases'</p> <p><i>EXCLUDE C:</i> Do not exclude studies based on their terms used to define 'linguistic units larger than one word'</p>	<p>This research seeks to assess the effectiveness of providing early foreign language learners with MWU input in teaching and learning settings. Thus, studies where no intervention is reported, where no emphasis is put on MWUs in the input, or—at least—where no correlations between MWU input and other linguistic/outcome variables are reported are not applicable since they provide no information as to the usefulness of MWU input in early foreign language learning settings</p> <p>MWUs are difficult to define linguistic phenomena, and much scientific work has slightly different views as to what constitutes MWUs. Part of the current research is to map out how different researchers operationalise MWUs in their work and why they do so, i.e., what are the expected learning effects (e.g., productivity, idiomaticity)</p> <p>A synthesis of empirical findings in this field of literature is impossible without the reporting and evaluation of concrete data</p>
Outcomes	<p><i>INCLUDE A:</i> Primary research studies reporting any measure of MWU input effectiveness, including but not limited to language outcomes (e.g., vocabulary uptake, grammatical skills), productivity outcomes (e.g., enhanced communicative abilities), or segmentation-/abstraction-/generalisation-ability outcomes. Include studies reporting either quantitative or qualitative outcomes</p> <p><i>EXCLUDE A:</i> Systematic reviews and studies that provide narrative evaluation of an educational program but provide no measures of MWU input effectiveness</p> <p><i>INCLUDE B:</i> All types of study design</p> <p><i>EXCLUDE B:</i> Do not exclude studies based on the research design</p>	<p>Research on this topic is scarce, and the exclusion of any one study design may provide an even narrower view of the research in this area</p>

TABLE 1 (Continued)

Item	Inclusion criteria	Rationale
Setting	Include all types of instructed settings, including but not limited to schools, after-school language clubs, psycholinguistic laboratories	
Publication status	Do not exclude studies based on publication status. Include grey literature	This paper seeks to offset potential publication bias by including a wider range of research, including grey literature
Language of publication	Do not exclude studies based on the language of publication	Limiting to studies written in English may result in a systematic neglect of a certain body of research

TABLE 2 List of databases.

Discipline	Database		
	English	German	French
Education	ProQuest Education Collection (including ERIC), British Education Index EBSCO	Fachportal Pädagogik	n/a
Linguistics	ProQuest Linguistics Collection (including LLBA)	n/a	n/a
Psychology	PsychInfo	n/a	n/a
Multidisciplinary	Web of Science, Scopus	Humboldt University Berlin (university library catalogue)	SUDOC, Pascal-Francis
Grey literature	ProQuest Dissertations & Theses Global	n/a	n/a

French

After consultation with French colleagues, *Pascal-Francis*, *SUDOC* (Système Universitaire de Documentation), and [Theses.fr](https://theses.fr) were initially considered suitable. Following pilot searches, [Theses.fr](https://theses.fr) was subsequently excluded because it did not process Boolean strings. Instead, we used *Pascal-Francis*, a database in ‘exact, human and social sciences’ run by the library of the Centre National de la Recherche Scientifique, and *SUDOC*, a meta-catalogue of French university libraries run by the Agence Bibliographique de l'Enseignement Supérieur.

In all languages, we conducted further searches using ‘citation-chaining’ (or ‘snowball searching’) which is considered best practice (Boland et al., 2017). In addition, forward citation searches were conducted in *Web of Science*.

Search strategy

To find all results available while avoiding unmanageable search outputs, we balanced sensitivity and specificity in search strings (Brunton et al., 2017). The initial string was created with the support of our department's librarian. The terms describing the concept ‘multi-word units’ in relevant literature were of concern because research provides no uniform conceptualisation. To cast a wide net, we used Christiansen and Arnon's (2017) list of 18 terms as reference for a search field specifying the language input participants received during studies. Other terms from relevant literature were also added (cf. Council of Europe, 2001; Wray, 2002).

As this review focuses on young learners aged 5–12, we added a ‘target participants’ search category. Multiple labels were included to represent the diverse terminology of formal teaching contexts (e.g., primary/elementary/junior school). Following pilot scoping searches on *ProQuest Education*, the category ‘participants’ was assigned to the search frame ABSTRACT. Assigning the category to ALL FIELDS resulted in too many results because terms like ‘primary school’ appear in too many publications. The search category specifying the target object (i.e., L2) was assigned to TITLE to obtain a manageable number of results. In addition, since MWUs are this review's key characteristic, we assigned NOFT (i.e., all text except full text) to the category specifying ‘MWU’. All translated linguistic terms were double-checked in pertinent German/French publications. We used Boolean operators such as truncation in the search strings. Example strings are provided in [Table 3](#).

Selection process

Following duplicate deletion, the first author screened each title and abstract, excluding studies that unambiguously violated one or more inclusion criteria. Being blind to the first author's decisions, the second author independently screened titles and abstracts of a randomly selected 10% sample of all records (87 × English; 51 × German; 33 × French). Afterwards, results were unblinded and the two authors discussed every conflict until a conclusion was reached (conflicts: 6 × English, $\kappa=0.93$; 4 × German, $\kappa=0.90$; 0 × French, $\kappa=1$). Having retrieved the records marked for full-text screening, the first author screened all full-texts and excluded studies that violated one or more inclusion criteria. Given the high kappa value from the first round of peer-screening, another round was deemed unnecessary. At this stage, difficult inclusion decisions were discussed among all authors.

Data collection process

Prior to the final searches, a data extraction form was created based on Boland et al. (2017) and the Cochrane Good Practice Guide (Cochrane Effective Practice and Organization of Care, 2017) (completed forms in Data S1). The form included all essential PICO items (Participants, Intervention, Comparison, and Outcomes; Petticrew & Roberts, 2006). Another section was added covering the type of MWU operationalisation used in the studies. In one case where data was reported insufficiently, the authors were successfully contacted for additional information.

One included study reported extensive qualitative evidence (~300 pages), rendering the original data extraction form unsuitable. Therefore, following approaches in other syntheses (e.g., Carlsen et al., 2007), the qualitative report was read repeatedly to identify key concepts. Based on those, an additional qualitative data extraction sheet was created (completed form in Data S2) which—corresponding to relevant Cochrane guidance—was aimed at synthesising ‘qualitative evidence within a stand-alone, but complementary, qualitative review to address questions on aspects other than effectiveness’ (Noyes & Lewin, 2011: 8). Given the large volume of qualitative data, the outcomes most relevant to the review's research purpose were selected as recommended by Noyes and Lewin (2011) and Page et al.'s (2021) latest PRISMA guidance. This process resulted in an analysis approach extracting data in the following categories, closely following the original author's qualitative data analysis:

- a. Which slot-filling/insertion word(s) was/were used?
- b. Does the transcript example represent paradigmatic or syntagmatic variations of formulaic sequences?
- c. Are segmentations of formulaic sequences observed?
- d. Did the student self-actively recombine formulaic sequences?
- e. Could the student realise their personal communicative goals?

This extraction approach focused on any MWUs and slot-fillers taught during the intervention, tracing them back to the individual transcription examples. The extraction ignored data irrelevant to the review's purpose, such as descriptions of students' degree of extroversion.

Following the first author's data extraction, the second author independently extracted data for one of the two included studies, resolving discrepancies through discussion. Since the other included work was an >800-page mixed-methods study written in German, the second author was only able to independently extract the quantitative data. An independent

TABLE 3 Example search strings.

Database		Search string
English	ProQuest education	ti(efl OR esl OR tefl OR mfl OR tesol OR tesl OR "second language*" OR "foreign language*" OR "artificial language*" OR english* OR fl OR L2 OR SLA) AND noft(multiword* OR multi-word* OR multiunit* OR frame OR frames OR idiom* OR multi-unit* OR prefabricated* OR "pre-fabricated*" OR phras* OR collocation* OR formulaic* OR "fixed expression*" OR "semi-fixed*" OR listeme* OR mwu OR mwe OR formula* OR chunk* OR routine* OR "sentence pattern*") AND ab("primary school*" OR "elementary school*" OR child* OR kids OR "young learner*" OR "grade school*" OR "infant school*" OR "early year*" OR "elementary grade*" OR kindergar?En OR "junior school*") AND ab(productivity OR ability OR vocabulary OR competence* OR "linguistic resource*" OR skill* OR "language knowledge" OR proficiency OR creativity OR "language uptake" OR development)
German	Fachportal Pädagogik	((((Abstract: „Kuenstliche Sprache "OR DAZ oder DAZ oder "DEUTSCH ALS ZWEITSPRACHE" oder DAF oder "DEUTSCH ALS FREMDSPRACHE" oder L2 oder SLA oder TEFL oder TESOL oder TESL oder ENGLISCH* oder FRANZOESISCH* oder FREMDSPRACH* oder ZWEITSPRACHE oder ZWEITSPRACHERWERB oder ZWEISPRACHIG*) und (Abstract: KONSTRUKTION oder FORMEL* oder WENDUNG oder PATTERN oder SATZFORMEL oder KOLLOKATION* oder IMITATION oder REPRODUKTION oder IMITIEREN oder AUTHENTISCH oder PHRASE oder AUTOMATISIEREN oder SCHEMA oder "VERBALE STEREOTYPE" oder CHUNK)) und (Abstract: SCHUELER* oder GRUNDSCHUL* oder JUNG* oder LERN* oder KIND* oder FRUEH* oder PRIMARSTUFE)) und (Freitext: PRODUKTIVITAET oder "KOMMUNIKATIVE FAEHIGKEIT" oder VOKABEL* oder SPRACHKOMPETENZ oder KOMPETENZ oder "LINGUISTISCHE RESSOURCE" oder KREATIVITAET oder WORTSCHATZ oder LEISTUNG)) und (Datenquelle: "FIS Bildung" oder "Library of Congress" oder "Casalini libri" oder ERIC oder "EBSCOhost ebooks" oder "BBF 1945–1993" oder "Online Contents" oder BASE)
French	Pascal-Francis	ti.*:(LVE OR L2 OR SLA OR TEFL OR TESOL OR TESL OR "non-natifs" OR "langue vivante" OR "langue moderne" OR "langue étrangère" OR "deuxième langue" OR "langue seconde" OR FLS OR FLE OR enseigne* OR anglais* OR allemand* OR langue* OR "français langue seconde") AND (collocation* OR congloméré* OR forme* OR lexème OR locution* OR métaphore* OR "mot composé" OR phrase* OR syntagme* OR "unité phraséologique" OR formul* OR "unité polylexicale" OR "expression polylexicale" OR séquence* OR SF OR phraséologi* OR "expression figée" OR idom*) AND ("École primaire" OR "école maternelle" OR "enseignement primaire" OR "école élémentaire" OR élève* OR apprenant* OR enfant* OR jeune*) AND (vocabulaire* OR productivité OR parole* OR compétence* OR créativité OR appropriation* OR acquisition* OR gramma*)

qualitative data extraction accounting for more than 300 pages of the study was deemed unfeasible because none of this review's other authors is a native German speaker. To establish the most objective and coherent data extraction process possible, the first and second author discussed the first author's approach to qualitative data extraction in detail. This process included the second author selecting sections of qualitative data at random and checking whether information was adequately represented on the data extraction form. In all cases, the data was documented comprehensively and thoroughly.

Risk of bias

Since research on MWUs in primary school L2 teaching contexts is scarce, we included all available research designs in the selection process, following Slavin's (1986) best-evidence synthesis, which does not discriminate against any research designs during study selection.

The Mixed Methods Appraisal Tool (MMAT; Hong et al., [n.d.](#); Pluye et al., 2009) allowing researchers to appraise research quality in five methodology categories (qualitative, randomised controlled trials, non-randomised, quantitative descriptive, mixed methods) was employed as the critical appraisal tool. This instrument has been updated and revised repeatedly (Hong et al., 2019; Pace et al., 2012) and used in similar systematic reviews including quantitative and qualitative designs (Richter, 2021; Willis et al., 2019).

Usually, each study receives quality scores of 0–5 per assessment category, which amounts to a global quality score, facilitating cross-study comparisons. However, the presentation of global quality scores has been discouraged in the past because metric scores cannot represent a study's problematic elements (Hong et al., [n.d.](#)). Since the current review only includes two studies, the individual risk of bias assessments for each study are reported instead of global ratings.

The second author independently completed the MMAT form for the included study published in English. Interrater reliability was high ($\kappa=0.86$), and the sole discrepancy emerged from missing information that the original author had only provided via e-mail. Since the second included study was published in German and consisted of more than 800 pages it was difficult for the co-authors (non-native speakers of German) to complete the MMAT independently. Therefore, the first and second author (German L2 speaker) discussed the completed MMAT in detail, the first author providing evidence for each rating in the original report (including translations). Agreement was reached on all ratings.

Synthesis methods

Since the number of included studies was small ($n=2$) and they differed considerably regarding their methodology (quantitative vs. mixed-methods) and design (i.e., intervention vs. corpus analyses and participatory action research), there was not enough comparable quantitative data for statistical meta-analyses. Instead, following Thomas et al. (2004), we triangulated qualitative and quantitative data in a narrative synthesis of study quality and findings.

RESULTS

Results of search and selection processes are presented in Figure 1. Of the 29 English reports selected for full-text screening, 7 were published in Korean (Chae & Kim, 2019; Jeon & Kim, 2018; Jung & Shin, 2013; Kim & Kang, 2005; Kim & Lee, 2009; Lee & Jeong, 2010;

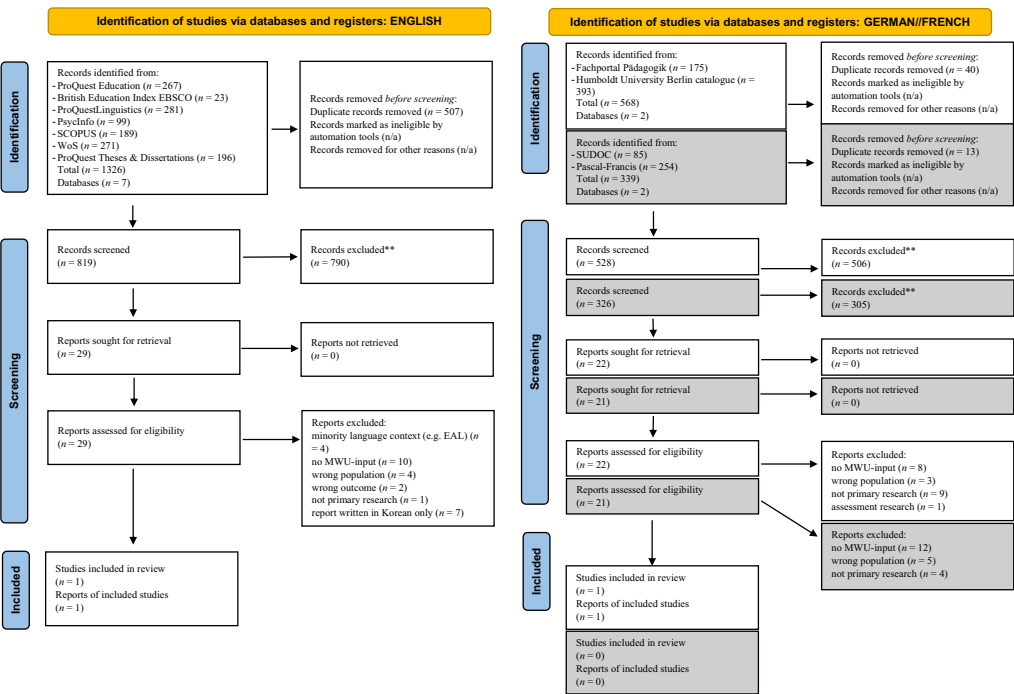


FIGURE 1 Selection process. Left side: English. Right side: German/French.

Lim & Lee, 2012). Requests to each paper's authors for an English version were unfruitful. Numerous translation attempts showed that the works' reference lists included additional potentially relevant works. Eventually, all seven papers were excluded because no member of the authors' research groups was a Korean-speaking linguistics expert, and translations were considered excessively labour-intensive and unreliable as they could not be evaluated by proficient Korean speakers.

Included studies

General study characteristics of the two included studies are provided in Table 4, followed by narrative summaries of the design, context and findings of each included study.

Balcı and Çakır (2012)

Balcı and Çakır (2012) investigated whether teaching new vocabulary through collocations results in better learning than teaching through 'classical' techniques (original report's terminology) by conducting a teaching intervention study with two 7th grade classes at a Turkish state primary school. The participants' age ranged from 12 to 14 years. We included the study because it is conducted in primary school and the age range is within the scope of the review protocol's age range (5–12). Both the control and intervention group received 4 h of English classes per week over a 6-week period. Whilst the control group learned vocabulary through techniques including direct translation and synonym/antonym tasks, the intervention group was presented with two reading passages per week (2 h per passage) including the same vocabulary items implemented in collocation contexts. After students had read the

TABLE 4 General characteristics of included studies.

Study	Balcı and Çakır (2012)	Kostka (2020)
Citation	Balcı and Çakır (2012)	Kostka (2020). <i>Produktives Sprechen im Englischunterricht der Grundschule: Eine empirische Studie zur Bedeutung formelhafter Sequenzen</i> . Doctoral dissertation. Giessen: University Library Publications
Publication Status	Journal article	Published PhD thesis
Study Design	Intervention	Observational (<i>Prospective study</i> ; cf. Petticrew & Roberts, 2006, p.134)
Data type	Quantitative	Quantitative/qualitative
Country	Turkey	Germany
Sample Size	$n=59$	Quantitative: $n=18$ Qualitative: $n=7$
Recruitment	State primary school in Konya	Three state primary schools near Frankfurt am Main
Study duration	6 weeks	1 school year (October–July)
Non-MWU control group?	Yes	No
General outcomes	Vocabulary uptake, vocabulary retention	Productive speaking skills (e.g., paradigmatic variations of patterns)
Specific outcome measures	Non-standardised vocabulary tests (multiple choice and gap-filling)	Quantitative: token frequency of formulaic sequences (patterns/routines) in speech corpus; token/type frequency of patterns in speech corpus; token/type frequency of individual patterns in speech corpus Qualitative: communicative-functional analysis of transcription examples from speech corpus

passages, the included collocations were discussed in class and presented in collocation networks on the blackboard. In both groups, the authors conducted a general vocabulary-based multiple-choice pre- and post-test on English proficiency, plus gap-filling vocabulary tests on the specific vocabulary used in the intervention at the end of each week (six in total). One week after the study, they conducted an additional retention test with both groups containing some of the gap-filling exercises from the previous weekly vocabulary tests. All tests were non-standardised.

The experimental group outscored the control in the last weekly test ($p<0.01$, Cohen's $d=0.87$) and in both the post-test and retention test (post-test: $p<0.01$, Cohen's $d=1.16$; retention test: $p<0.01$; Cohen's $d=0.78$). Since there were no such significant differences between the groups at pre-test, or in the first five weekly tests, the authors conclude that post-training differences are due to the intervention. Unfortunately, this conclusion is too strong, since a significance test on pre-test difference cannot provide evidence for a lack of differences between the groups at that time point (a non-significant p -value does not show whether there is evidence for the null-hypothesis, though it is routinely misinterpreted in this way—cf. Dienes, 2008, 2014; Lakens, 2017). Similarly, where the authors do look at pre- to post-test differences (uptake over time) they look at this for each group separately, finding within-group (week 1 vs. week 6) significant improvement in test scores for the experimental group ($p<0.01$, Cohen's d =not reported) but not the control group. Again, however, inferring from this that gains in the experimental group are larger than those in the control

is to overinterpret the null result in the control condition. Without a direct comparison of the gains in the two groups (e.g., using a time-by-group interaction) it is difficult to draw strong conclusions from this data. Thus, though the patterns are certainly suggestive, the authors' conclusion that the students benefitted from learning new vocabulary through collocations, and that this effect specifically showed up after 5 weeks of training, must be treated with some caution.

Kostka (2020)

Kostka (2020) investigated the role of formulaic sequences in German primary students' productive English skills. Supported by three participating primary school English teachers, Kostka designed and initiated a planned, controlled, gradual reconfiguration (original report's terminology) of the English classes of three Year 3 classes (8- to 9-year-olds) in schools around Frankfurt am Main, Germany, over a period of one school year. Her work was not intended to conform to classic intervention designs as no control group was involved. The teaching reconfiguration consisted of increased teaching foci on a set of formulaic sequences subclassified into 'patterns' (with open slots) such as *My favourite _ is _* and 'routines' (fixed constructions) such as *What's your name?* Considering previous MWU research, Kostka anticipated that this input focus would improve students' productive skills over time. The reconfiguration was divided into three teaching phases and an end-of-year assessment phase. During the school year, Kostka collected students' oral speech data from video recordings of classroom interactions and of targeted end-of-year assessment situations. While Kostka transcribed and analysed quantitatively the speech data of 18 students, she analysed qualitatively the speech data of 7 students. In both cases, the selected students covered a range from weak to strong students.

Overall, Kostka reports that formulaic sequences accounted for about 50% of all articulated words. Based on token frequencies, her quantitative corpus analyses revealed that irrespective of their proficiency level, students used *patterns* more frequently than *routines* in all three teaching phases, the assessment phase and the entire year's corpus. The overall use of *formulaic sequences* decreased significantly from teaching phase one to the assessment phase ($p < 0.0001$). According to Kostka, this decrease supports the assumption of a parallel running process of students' increasing segmentation and analysis abilities. The conclusion that students' productive speaking skills improved over the year is supported by that fact that Giraud's Index (a measure of lexical diversity) of the 'content areas' that students drew from to fill slots in *patterns* increased steadily over the year. That is, the quantitative data suggest that students gradually inserted words from an increasing variety of contexts into empty slots of *patterns*.

In her qualitative analyses, Kostka conducted communicative-functional analyses of transcribed speech examples of seven students' productive-dialogic speaking over the course of all three teaching phases and the assessment phase. In her analyses, the author considered individual student characteristics (data collected during classroom observations) both on a personal level (e.g., extroversion) and a subject-related level (e.g., language awareness). Overall, the qualitative analyses revealed that all seven students could self-actively implement formulaic sequences into spontaneous dialogues to reach their personal communication goals. To varying degrees, both weak and strong students could apply learned lexical knowledge to new linguistic contexts. At the beginning of the year, weaker students relied on a smaller set of formulaic sequences whereas stronger students integrated a comparatively large spectrum of routines and patterns into their dialogues. Paradigmatic variations of patterns played a central role in the gradually increasing productive use of formulaic sequences throughout the school year. This qualitative observation corresponds to the quantitative

findings. However, weaker students relied on patterns longer than strong students, who started showing an increasing ability to segment and analyse patterns and reconfigure individual segments earlier. Syntagmatic variations also occurred more often among strong students than weak students. Regarding the stronger students, one exercise format in the assessment phase required them to dismantle the learned routines and patterns, initiating linguistic transfers of structural knowledge to new contexts, thereby increasing their segmentation abilities.

Risk of bias assessments for included studies

Detailed risk of bias assessments are provided in Data S3. We consider Kostka's (2020) overall level of risk of bias to be low, confounded only by the non-consideration of students' language biographies. Balcı and Çakır's (2012) overall level of risk of bias we consider high, for example because all employed measurements were non-standardised.

Summary statistics

Summary statistics for Balcı and Çakır (2012) are provided in Data S4. Kostka (2020) used no classic intervention design (i.e., no control group), therefore, summary statistics for each group and effect estimates cannot be reported. Descriptive statistics relevant to this review's research questions and corresponding within-group differences of mean difference between the individual teaching/assessment phases are provided in Data S5. Summaries of Kostka's (2020) qualitative findings are provided in Data S2 and S6.

Results of syntheses

RQ1: What is the extent of original research investigating the role of MWUs in early FL learning and teaching contexts?

The empirical research output on MWU's role in early L2 learning is scarce. While many works that were excluded in the screening phase provided theoretical accounts of MWU's important role in language education, and some works reported empirical work from other age groups, only two works were retrieved that report relevant empirical data from primary school contexts. Seven works were excluded because they were only available in Korean. They could not be excluded based on their English abstracts and therefore may well have been eligible for inclusion in the current review.

RQ2: What is the extent of the evidence on the effectiveness of learning from/teaching MWUs in early FL contexts?

2a. What impact does MWU input have on young learners' grammatical and vocabulary skills and knowledge in the target language?

Since only two studies were identified in our review, the extent to which these studies can address our research questions is limited.

Only Balcı and Çakır (2012) provide test data from vocabulary tests. However, since the authors do not report time-by-group interactions it is difficult to discern whether there was indeed a between-group difference in pre- to post-test gains. We cannot say whether the

intervention influenced vocabulary uptake, although this is consistent with the descriptive statistics.

Direct measures of grammatical knowledge are not reported in the included studies. Kostka (2020) provides indirect accounts of students' grammatical skills as she describes a gradual development of rudimentary top-down processes by some strong students (e.g., student S1a). Which exact grammatical patterns those students have acquired remains unclear.

2b. What impact does MWU input have on young learners' communication abilities (e.g., quantity and quality of spoken output; receptive/productive vocabulary knowledge)?

Balcı and Çakır (2012) do not report data on their learners' communication abilities as their vocabulary tests measure single-word knowledge. Kostka (2020) finds MWU use has positive effects on students' communication abilities. Given the small number of participants and the lack of control group, Kostka (2020) only tentatively ascribes the students' linguistic development to the implemented teaching changes. Nonetheless, she takes into consideration the possibility of a relationship between the teaching focus on formulaic sequences and the findings from the corpus data. In her study, MWU input gave students relatively strong communicative agency, enabling them to realise personal communication goals in their first year of L2 study. The reported quantitative and qualitative findings complement one another in this respect. Although the corpus data showed that the overall number of formulaic sequences gradually decreased over time, Giraud's Index of 'content areas' and the qualitative analyses indicate that the decrease was due to an uptake of segmentation and analysis abilities. Based on her analyses, Kostka concludes that the variety of communication goals the students managed to realise throughout the year via (minimal) formal-linguistic variations of learned formulaic sequences indicates that productive speaking can already be initiated in the first year of primary school L2 classes. Importantly, Kostka reports differences between weak and strong students regarding the impact of MWU input on communication abilities. Whereas weaker students rely longer on formulaic sequences, stronger students start developing segmentation abilities and applying abstracted pattern-knowledge productively earlier.

With regards to the general result of this systematic review, note that Balcı and Çakır (2012) received low quality ratings, weakening the reliability of their results, and Kostka (2020) did not include a control group in her study. Rightly, she emphasises that her findings cannot unambiguously be traced back to the implemented teaching changes. Although the reported findings are promising, considering the paucity of available evidence, this systematic review cannot report trustworthy evidence of the effectiveness of teaching MWUs to young L2 learners.

DISCUSSION

Available research

The review's results demonstrate that research into the effectiveness of MWU input in early L2 teaching is scarce and available evidence is inconclusive. These results are striking because MWUs are considered vital for L2 development in linguistics, psycholinguistics, cognitive linguistics and language pedagogy (Siyanova-Chanturia, 2017). From a pedagogical perspective, the lack of research is startling as EFL is part of many primary school curricula worldwide, and in Europe, MWUs are already an integral part of said curricula (see the Introduction). Scientific evidence supporting the effectiveness of MWU input in early

L2 teaching is crucial to (a) legitimate and reinforce the importance of MWUs in curricula, teaching materials and teacher education, and (b) provide important groundwork facilitating the implementation of targeted changes in curricula and teaching to further improve learning outcomes.

The review may be biased due to the inaccessibility of seven Korean papers we excluded because English versions were unavailable. The inaccessibility limits the review because the methodologies and results reported in the English abstracts were promising, indicating that MWU input improved students' L2 abilities (Lee & Jeong, 2010), that control groups were used (Lim & Lee, 2012), and that interventions positively impacted students' language learning motivation and self-esteem (Jung & Shin, 2013). Unfortunately, with English being the lingua franca of scientific publishing, these papers are difficult to retrieve, rendering their contents mostly inaccessible. However, even if each of the seven Korean language studies had been included, that would still amount to only nine available pieces of original research addressing what we believe to be an important theoretical and practical question. Therefore, despite this potential bias our conclusion that there is a vanishingly small amount of research in this area still holds.

Importantly, as stressed by Gray (2021), identifying only two studies in this review is nonetheless a valuable finding for current L2 research because we can now be confident that there is little relevant research available; researching whether MWUs in curricula and teaching have beneficial effects on young learners' L2 development is a much-needed area of future inquiry. In addition, another identified area for much-needed work concerns the conceptualisation of MWUs in classroom-based language research. In our review, this theoretical issue cannot sufficiently be considered because no consistent picture appears with the limited amount of research available. However, the more research is conducted in this area, the more pressing the issue of consistent conceptualisations of 'MWU' will become.

In addition to the scarcity of research is the issue of the quality of available research. On the one hand, Kostka (2020) reports meticulous work, but her findings cannot be tied to the teaching because a control group was lacking. On the other hand, Balcı and Çakır's (2012) study includes a control, but there are other limitations such as non-standardised testing. Considering the limited and mixed evidence-base, the following discussion of the second research question should therefore be regarded as tentative.

Effectiveness of MWU input

Balcı and Çakır's (2012) results seem to suggest that the teaching intervention was successful, yet their statistical analyses were inconclusive. Nonetheless, their data evoke consideration of the appropriateness of different measurements in MWU research contexts. On the one hand, Balcı and Çakır (2012) rightly used 'traditional' vocabulary measurements such as single-word gap-filling exercises corresponding to their research aim; namely, to investigate single-word vocabulary uptake. On the other hand, if the role of MWUs in language development in future work is considered in constructivist learning frameworks such as UBL, then researchers should introduce different types of measurement. That is, future research aiming at communication-related dependent variables such as productivity would benefit from the use of measurements that can capture MWU knowledge. In fact, researchers have proposed ways of measuring MWU knowledge, such as using mutual information (MI) scores in speech output (Polio & Yoon, 2020). Others have introduced tests such as the multi-word phrase test (MPT) measuring children's multi-word vocabulary knowledge based on *verb + object* phrase knowledge (Smith & Murphy, 2015). Since the purpose of early MWU input is (a) the provision of early communicative agency and (b) the catalysis of generalisation and abstraction processes, future research must avoid unreliable measurements, which

might misrepresent young L2 learners' language knowledge and blur our understanding of the effectiveness of MWU input.

In general, instead of specifically asking about learners' vocabulary and grammar gains, perhaps more relevant questions concerning MWU input are about learners' communication abilities as in Kostka (2020). Such inquiries correspond to main goals of current primary school L2 teaching to make children active language users and to establish early communication ability (see the Introduction). Beginning L2 primary learners rely heavily on MWUs in the first stages of their development as such linguistic units are their first and only means for successful L2 communication (Diehr & Polte, 2009). Having immediate linguistic agency to realise personal communication goals increases (or at least maintains) children's generally high language-learning motivation (Hempel, 2016; Lenzing & Roos, 2012). And although MWUs are important linguistic agents for weak learners to communicate in the first place, they are catalysts for stronger learners to abstract patterns and advance to free productions (Kahl & Knebler, 1996; Kostka, 2020; Sambanis, 2007). Congruously, Kostka (2020) did not simply administer vocabulary tests but analysed her qualitative data relative to the students' ability to realise personal communicative goals. Supported by her quantitative findings, the qualitative work confirms what other research (Kahl & Knebler, 1996; Myles et al., 1999; Sambanis, 2007) has suggested and what curricula (e.g. KM-BW, 2016) have expected, namely that students rely heavily on MWUs to communicate successfully. In addition, her data show that weaker learners relied longer on formulaic sequences than stronger learners, who started dismantling phrases and using abstracted structural knowledge productively earlier. Concerning communication abilities, both proficiency levels benefitted from the use of MWUs enabling them to realise basic communicative goals or to fill slots in MWUs and rearrange and combine patterns to realise advanced goals. Therefore, following Kostka's approach, measurements of the effectiveness of MWU input should be regarded against the background of children's linguistic generativity instead of solely concentrating on their vocabulary and grammar uptake. Although such an approach to the measurement of communication abilities appears promising for future work, the lack of a control group in Kostka (2020) makes it impossible to discern to what extent the uptake in students' communication ability was a direct result of the input's MWU focus.

Regarding communication ability, Kostka's data suggest that the type of task students complete influences the amount of their productive speaking. For example, in the unfocused task *5 min chat*, students discussed topics of their choice without being provided with example formulaic sequences. Most students were able to reactivate learned formulaic sequences and recombine them according to the respective communicative contexts. Some students also managed to 'unhinge' formulaic sequences from their original context and use them for personal communication goals. In contrast, in focused tasks such as *information gap activities*, students were provided with predetermined communicative contexts, forcing them to modify the learned formulaic sequences accordingly (e.g., change learned first and second person singular to third person singular). In many cases, this modification led to the segmentation of the initial formulaic sequence and eventually to free constructions based on that segmentation information. That is, this type of exercise seems to have led students to dismantle learned formulaic sequences, and extract and use underlying structural information productively. This observed influence of exercise type on linguistic development processes is important for future measurements of MWU knowledge and, more generally speaking, communication abilities, because it indicates that linguistic development processes are inevitably context dependent, as suggested by UBL (Eskildsen, 2009). As demonstrated in previous work (e.g., Dahl, 2015), Kostka (2020) demonstrates that authentic and relevant contexts can stimulate young students' communicative scopes. Having partly predetermined contexts motivates students to 'play' with the learned MWUs and seems to create a fruitful catalyst for self-initiated abstraction and generalisation processes (Kostka, 2020).

Kostka's (2020) data show that cognitive processes such as rule abstraction are slow and rely on repetitive input and on context conditions, such as exercise type (cf. Eskildsen, 2009; Nattinger & DeCarrico, 1992; Wulff, 2018). In this context, researchers have repeatedly warned about establishing realistic L2 outcomes among primary school learners (Jäger, 2012). The time spent in L2 classrooms varies widely across contexts and for many children is less than 1 h per week. Language teaching often plays a subordinate role in schools' curricula and in some countries, such as Germany, L2 teaching only starts in Year 3. Therefore, young learners will only ever be able to achieve rudimentary communication skills (Jäger, 2012). Although results from longitudinal large-scale L2 evaluations in German primary schools have indicated that learners benefit from MWUs and gradually start breaking them down, realistically, communication skills by the end of Year 4 are still rudimentary among most students (Hempel, 2016). For example, Engel et al. (2009) and Pienemann et al. (2006) report that despite high motivation levels among learners, they were often not able to realise their communicative goals fully and by the end of Year 4 many students, especially weaker ones, were stuck on a reproductive and imitative plateau. This corresponds to Kostka's (2020) finding that only stronger students were able to segment patterns and produce speech freely. Therefore, considering limited teaching time, lack of appropriate teacher L2 education (Holmes & Myles, 2019), and the observation that MWU input can only help so much, it appears sensible to support MWU input with corresponding rule-input and training of metalinguistic awareness. As demonstrated in Tode's (2003) study on copula-acquisition among beginning Japanese EFL learners, the 'chunk-rule continuum' in instructed settings does not only develop based on inductive rule-abstraction but also 'normal' deductive rule-input is important. Similarly, Edmondson (1995) argued that grammar and lexis must be purposefully 'connected' in the L2 curriculum, and Wulff (2018) summarises that 'inductive construction learning ... complements deductive, rule-based learning processes' (p. 19). Thus, although young learners in fact demonstrate developing abstraction and slot-filling abilities (Kostka, 2020; Sambanis, 2007), researchers, teachers and other stakeholders should remain realistic about L2 attainment goals at the end of primary school, and keep in mind that MWU input is no panacea in teaching young L2 learners, making other types of input relevant too.

CONCLUSION

This review sought to determine the scope and extent of original research investigating the effectiveness of MWU input on young learners' L2 development. It demonstrates that evidence in this field is scarce, and that consequently we know little about whether and how MWU input in primary schools is beneficial for students' L2 development. However, we have seen that Kostka's (2020) extensive qualitative data underlines MWU's role in providing young learners with early communicative agency, which is indispensable for students' motivation and self-confidence (Lenzing & Roos, 2012). Although the included studies indicate that current L2 curricula and teaching are on the right track considering constructivist L2 learning theories, researchers have much work ahead to better understand the impact of MWU input in primary school L2 teaching on linguistic variables such as segmentation skills. Fundamentally, the scant review results reinforce calls for more intervention studies in language education (Chalmers & Murphy, 2021). Without question, high-powered large-scale randomised controlled trials are difficult to fund, organise, and conduct. Yet, even small-scale quasi-experimental intervention studies are valuable signposts for researchers and educators alike. Multiple small-scale studies with similar patterns of results lay necessary groundwork for potential larger interventions, and they confirm current teaching practices. As starting points for such future work, this review demonstrates that factors such as learner proficiency and exercise type are promising avenues

for investigation. Furthermore, our findings re-evolve notions that the appropriateness of the assessment instrument in MWU studies is crucial (Smith & Murphy, 2015). By avoiding the limitations of the studies discussed here, future work can ideally confirm that MWUs are indeed catalysts for young learners' L2 development in instructed settings.

CONFLICT OF INTEREST STATEMENT

The authors report no conflict(s) of interest.

FUNDING INFORMATION

Johannes Schulz is funded by the Department of Education, University of Oxford.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Johannes Schulz  <https://orcid.org/0000-0003-3769-2461>

REFERENCES

- (KM-BW) Kultusministerium Baden-Württemberg. (2004). *Bildungsplan 2004. Grundschule*. http://www.bildungsplaene-bw.de/site/bildungsplan/get/documents_E-1204433591/lsbw/Bildungsplaene/Bildungsplaene-2004/Bildungsstandards/Grundschule_Bildungsplan_Gesamt.pdf
- (KM-BW) Kultusministerium Baden-Württemberg. (2016). *Bildungsplan der Grundschule. Englisch (ab Klasse 3/4)*. https://www.bildungsplaene-bw.de/site/bildungsplan/get/documents/lsbw/export-pdf/depot-pdf/ALLG/BP2016BW_ALLG_GS_E34.pdf
- Aguado, K. (2002). Formelhafte Sequenzen und ihre Funktionen für den L2- Erwerb. *Zeitschrift für Angewandte Linguistik*, 37, 27–49.
- Anron, I., & Christiansen, M. H. (2017). The role of multiword building blocks in explaining L1–L2 differences. *Topics in Cognitive Science*, 9(3), 621–636. <https://doi.org/10.1111/tops.12271>
- Anron, I., McCauley, S. M., & Christiansen, M. H. (2017). Digging up the building blocks of language: Age-of-acquisition effects for multiword phrases. *Journal of Memory and Language*, 92, 265–280. <https://doi.org/10.1016/j.jml.2016.07.004>
- Balci, O., & Çakır, A. (2012). Teaching vocabulary through collocations in EFL classes: The case of Turkey. *International Journal of Research Studies in Language Learning*, 1(1), 21–32.
- Bannard, C., & Lieven, E. (2012). Formulaic language in L1 acquisition. *Annual Review of Applied Linguistics*, 32, 3–16. <https://doi.org/10.1017/S0267190512000062>
- Behrens, H. (2009). Usage-based and emergentist approaches to language acquisition. *Linguistics*, 47(2), 383–411. <https://doi.org/10.1515/LING.2009.014>
- Bergelson, E., & Aslin, R. N. (2017). Nature and origins of the lexicon in 6-mo-olds. *Proceedings of the National Academy of Sciences of the United States of America*, 114(49), 12916–12921. <https://doi.org/10.1073/pnas.1712966114>
- BIG-Kreis. (2015). *Der Lernstand im Englischunterricht am Ende von Klasse 4. Ergebnisse der BIG- Studie*. Domino.
- Boers, F., Dang, T. C. T., & Strong, B. (2017). Comparing the effectiveness of phrase-focused exercises: A partial replication of Boers, Demecheleer, Coxhead, and Webb (2014). *Language Teaching Research*, 21(3), 362–380. <https://doi.org/10.1177/1362168816651464>
- Boers, F., Eyckmans, J., Kappel, J., Stengers, H., & Demecheleer, M. (2006). Formulaic sequences and perceived oral proficiency: Putting a lexical approach to the test. *Language Teaching Research*, 10(3), 245–261. <https://doi.org/10.1191/1362168806lr195oa>
- Boers, F., & Lindstromberg, S. (2012). Experimental and intervention studies on formulaic sequences in a second language. *Annual Review of Applied Linguistics*, 32, 83–110. <https://doi.org/10.1017/S0267190512000050>
- Boers, F., & Muñoz-Basols, J. (2021). Acquisition of idiomatic language in L2 Spanish. In J. Barcroft & J. Muñoz-Basols (Eds.), *Spanish vocabulary learning in meaning-oriented instruction* (pp. 62–88). Routledge.
- Boland, A., Cherry, M. G., & Dickson, R. (2017). *Doing a systematic review: A student's guide* (2nd ed.). SAGE.
- Bredenbröcker, M. (2018). *A corpus-based approach to English as a foreign language at primary school level: Collocations for early beginners*. Verlag Dr. Kovac.

- Brunton, G., Stansfield, C., Caird, J., & Thomas, J. (2017). Finding relevant studies. In D. Gough, S. Oliver, & J. Thomas (Eds.), *An introduction to systematic reviews* (2nd ed., pp. 93–122). SAGE.
- Carlsen, B., Glenton, C., & Pope, C. (2007). Thou shalt versus thou shalt not: A meta-synthesis of GPs' attitudes to clinical practice guidelines. *The British Journal of General Practice: The Journal of the Royal College of General Practitioners*, 57(545), 971–978. <https://doi.org/10.3399/096016407782604820>
- Chae, J., & Kim, H. (2019). The effects of receptive and productive collocation learning task types on elementary school Students' English vocabulary acquisition. *The Journal of Education*, 39(2), 399–422.
- Chalmers, H., & Murphy, V. (2021). Multilingual learners, linguistic pluralism and implications for education and research. In E. Macaro & R. Woore (Eds.), *Debates in second language education* (pp. 66–88). Routledge.
- Choi, S. (2017). Processing and learning of enhanced English collocations: An eye movement study. *Language Teaching Research*, 21(3), 403–426. <https://doi.org/10.1177/1362168816653271>
- Christiansen, M. H., & Arnon, I. (2017). More than words: The role of multiword sequences in language learning and use. *Topics in Cognitive Science*, 9(3), 542–551. <https://doi.org/10.1111/tops.12274>
- Cochrane Effective Practice and Organisation of Care (EPOC). (2017). *Data collection form*. https://epoc.cochrane.org/sites/epoc.cochrane.org/files/public/uploads/Resources-for-authors2017/good_practice_data_extraction_form.doc
- Große-Brauckmann, S. & Heidelberg, N. (Eds.). (2018). *Come in 3. Activity book*. Ernst Klett Verlag.
- Council of Europe. (2001). *Common European framework of reference for language learning, teaching, assessment*. Cambridge University Press.
- Council of Europe. (2018). *Common European framework of reference for languages. Learning, teaching, assessment. Companion volume with new descriptors*. <https://rm.coe.int/cefr-companion-volume-with-new-descriptors-2018/1680787989>
- Czarnecka, M. (2011). Formelhafte Sequenzen in der Erst- und Zweitsprache: Versuch einer Begriffsbestimmung aus psycholinguistischer Perspektive. *Germanica Wratislaviensia*, 133, 189–199.
- Dahl, A. (2015). Input and language competence in early-start foreign language classrooms. In T. Cadierno & S. W. Eskildsen (Eds.), *Usage-based perspectives on second language learning* (pp. 125–152). De Gruyter.
- Diehr, B. (2009). Young learners' use of English. Imitation or production? In T. Stewart (Ed.), *Insights on teaching speaking in TESOL* (pp. 53–66). TESOL.
- Diehr, B., & Polte, L. (2009). Zur Entwicklung diskursiver Fähigkeiten im Englischunterricht der Grundschule. Eine vergleichende Untersuchung von Sprechern des Englischen als Erst- und Fremdsprache. *Zeitschrift Für Fremdsprachenforschung*, 20(2), 147–174.
- Dienes, Z. (2008). *Understanding psychology as a science: An introduction to scientific and statistical inference*. Palgrave Macmillan.
- Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, 5, 781. <https://doi.org/10.3389/fpsyg.2014.00781>
- Edmondson, W. (1995). Wortschatzerwerb und Spracherwerb. In K.-R. Bausch, H. Christ, F. Königs, & H.-J. Krumm (Eds.), *Erwerb und Vermittlung von Wortschatz im Fremdsprachenunterricht* (pp. 55–62). Narr.
- Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24(2), 143–188. <https://doi.org/10.1017/S0272263102002024>
- Ellis, N. C. (2015). Cognitive and social aspects of learning from usage. In T. Cadierno & S. W. Eskildsen (Eds.), *Usage-based perspectives on second language learning* (pp. 49–73). De Gruyter Mouton.
- Ellis, N. C., & Ferreira-Junior, F. (2009). Construction learning as a function of frequency, frequency distribution, and function. *The Modern Language Journal*, 93(3), 370–385. <https://doi.org/10.1111/j.1540-4781.2009.00896.x>
- Ellis, N. C., & Wulff, S. (2020). Usage-based approaches to L2 acquisition. In B. VanPatten, G. D. Keating, & S. Wulff (Eds.), *Theories in second language acquisition* (3rd ed., pp. 63–82). Routledge.
- Enever, J. (2011). *ELLiE: Early language learning in Europe*. British Council.
- Engel, G., Groot-Wilken, B., & Thürmann, E. (2009). *Englisch in der Primarstufe – Chancen und Herausforderungen. Evaluation und Erfahrungen aus der Praxis*. Cornelsen.
- Eskildsen, S. W. (2009). Constructing another language—Usage-based linguistics in second language acquisition. *Applied Linguistics*, 30(3), 335–357. <https://doi.org/10.1093/applin/arn037>
- Eskildsen, S. W., & Cadierno, T. (2015). Advancing usage-based approaches to L2 studies. In T. Cadierno & S. W. Eskildsen (Eds.), *Usage-based perspectives on second language learning* (pp. 1–16). De Gruyter Mouton.
- Europarat. (2001). *Gemeinsamer europäischer Referenzrahmen für Sprachen. lernen, lehren, beurteilen*. Langenscheidt.
- Eyckmans, J., & Lindstromberg, S. (2017). The power of sound in L2 idiom learning. *Language Teaching Research*, 21(3), 341–361. <https://doi.org/10.1177/1362168816655831>

- Fillmore, L. W. (1976). *The second time around: Cognitive and social strategies in second language acquisition*. Doctoral dissertation, Stanford University, Stanford, CA. <https://www.proquest.com/docview/302810294?pq-origsite=gscholar&fromopenview=true>
- Graddol, D. (2006). *English next*. British Council.
- Gray, R. (2021). Empty systematic reviews: Identifying gaps in knowledge or a waste of time and effort? *Nurse Author & Editor*, 31(2), 42–44. <https://doi.org/10.1111/nae2.23>
- Hakuta, K. (1974). Prefabricated patterns and the emergence of structure in second language acquisition. *Language Learning*, 24(2), 287–297. <https://doi.org/10.1111/j.1467-1770.1974.tb00509.x>
- Hempel, M. (2016). Förderung produktiver Sprachkompetenzen – Sind Lehrwerke Teil des Problems oder Teil der Lösung? In H. Böttger & N. Schlüter (Eds.), *FFF – Fortschritte im Frühen Fremdsprachenlernen. Tagungsband zur 4. FFF-Konferenz 2014 in Leipzig* (pp. 124–133). Westermann.
- Holmes, B., & Myles, F. (2019). *White Paper: Primary languages policy in England – The way forward*. RiPL: www.ripl.uk/policy/. <https://www.ripl.uk/wp-content/uploads/2019/02/RIPL-White-Paper-Primary-Languages-Policy-in-England.pdf>
- Hong, Q. N., Pluye, P., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M.-P., Griffiths, F., Nicolau, B., O'Cathain, A., Rousseau, M.-C., & Vedel, I. (2019). Improving the content validity of the mixed methods appraisal tool: A modified e-Delphi study. *Journal of Clinical Epidemiology*, 111, 49–59. <https://doi.org/10.1016/j.jclinepi.2019.03.008>
- Hong, Q. N., Pluye, P., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M.-P., Griffiths, F., Nicolau, B., O'Cathain, A., Rousseau, M.-C., & Vedel, I. (n.d.). *Mixed methods appraisal tool (MMAT), version 2018*. Registration of Copyright (#1148552), Canadian Intellectual Property Office, Industry Canada. http://mixedmethodsappraisaltoolpublic.pbworks.com/w/file/attach/127916259/MMAT_2018_criteria-manual_2018-08-01_ENG.pdf
- Huang, J., & Hatch, E. M. (1978). A Chinese child's acquisition of English. In E. Hatch (Ed.), *Second language acquisition: A book of readings* (pp. 118–131). Rowley.
- Ibbotson, P. (2013). The scope of usage-based theory. *Frontiers in Psychology*, 4, 255. <https://doi.org/10.3389/fpsyg.2013.00255>
- Jäger, A. (2012). Die Förderung kommunikativer Fähigkeiten im Englischunterricht der Grundschule. In H. Böttger (Ed.), *Englisch. Didaktik für die Grundschule* (pp. 112–122). Cornelsen.
- Jeon, J.-H., & Kim, J.-R. (2018). Reflection on self-instruction in elementary English based on the frequency of new vocabulary by activity of COLT part 2. *Asia-Pacific Journal of Multimedia Services Convergent with Art, Humanities, and Sociology*, 8(9), 353–366.
- Jung, D.-W., & Shin, C.-O. (2013). The effects of chunk-focused learning on Korean elementary school Students' English skills and interest. *Primary English Education*, 19(1), 219–239.
- Kahl, P., & Knebler, U. (1996). *Englisch in der Grundschule—Und dann? Evaluation des Hamburger Schulversuchs Englisch ab Klasse 3*. Cornelsen.
- Kenyeres, A., & Kenyeres, E. (1938). Comment une petite hongroise de sept ans apprend le français. *Archives de Psychologie*, 26, 321–366.
- Kersten, S. (2015). Language development in young learners: The role of formulaic language. In J. Bland (Ed.), *Teaching English to young learners: Critical issues in language teaching with 3–12 year olds* (pp. 129–145). Bloomsbury Academic.
- Kim, G., & Lee, S. (2009). A study on vocabulary teaching through English stories. *English21*, 22(3), 157–186.
- Kim, H.-R., & Kang, W.-S. (2005). Using English cartoons for children's literacy development in the elementary school. *Primary English Education*, 11(1), 209–240.
- Kostka, N. (2020). Produktives Sprechen im Englischunterricht der Grundschule: Eine empirische Studie zur Bedeutung formelhafter Sequenzen. *Giessener Fremdsprachendidaktik:online*, (Vol. 15). Giessen University Library Publications. http://geb.uni-giessen.de/geb/frontdoor.php?source_opus=9579&la=de
- Lakens, D. (2017). Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, 8(4), 355–362. <https://doi.org/10.1177/1948550617697177>
- Lee, Y.-J., & Jeong, D.-B. (2010). The effect of collocation-based English vocabulary learning on Children's speaking abilities. *Studies in English Language and Literature*, 36(3), 293–334. <https://doi.org/10.21559/aellk.2010.36.3.016>
- Legutke, M., Müller-Hartmann, A., & Schocker-von Ditfurth, M. (2012). *Teaching English in the primary school*. Klett Lerntraining.
- Lenzing, A., & Roos, J. (2012). Die sprachliche Entwicklung und Ausdrucksmöglichkeiten von Grundschulinnen und Grundschulern im Englischunterricht. In M. Bär, A. Bonnet, H. Decke-Cornill, A. Grünwald, & A. Hu (Eds.), *Globalisierung, Migration, Fremdsprachenunterricht. Dokumentation zum 24. Kongress für Fremdsprachendidaktik der Deutschen Gesellschaft für Fremdsprachenforschung (DGFF) Hamburg* (pp. 207–220). Schneider.
- Lieven, E., Behrens, H., Speares, J., & Tomasello, M. (2003). Early syntactic creativity: A usage-based approach. *Journal of Child Language*, 30(2), 333–370. <https://doi.org/10.1017/S0305000903005592>

- Lieven, E., Salomo, D., & Tomasello, M. (2009). Two-year-old children's production of multiword utterances: A usage-based analysis. *Cognitive Linguistics*, 20(3), 481–507. <https://doi.org/10.1515/COGL.2009.022>
- Lim, E., & Lee, K. (2012). Elementary students' use of English collocation and its influence on written language learning and learning attitudes. *Journal of British & American Studies*, 26, 265–298.
- Mellow, J. D. (2006). The emergence of second language syntax: A case study of the acquisition of relative clauses. *Applied Linguistics*, 27(4), 645–670. <https://doi.org/10.1093/applin/aml031>
- Müller, T., Böttger, H., Schlüter, N., Kierepka, A., Börner, O., Legutke, M., Kronisch, I., & Lohmann, C. (2016). *Der Lernstand im Englischunterricht am Ende von Klasse 4: Erste Ergebnisse der BIG-Studie*. Domino.
- Myles, F., Mitchell, R., & Hooper, J. (1999). Interrogative chunks in French L2: A basis for creative construction? *Studies in Second Language Acquisition*, 21(1), 49–80. <https://doi.org/10.1017/S0272263199001023>
- Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. OUP.
- Noyes, J., & Lewin, S. (2011). Extracting qualitative evidence. In J. Noyes, A. Booth, K. Hannes, A. Harden, J. Harris, S. Lewin, & C. Lockwood (Eds.), *Supplementary guidance for inclusion of qualitative research in Cochrane systematic reviews of interventions*. Version 1 (updated august 2011). Cochrane Collaboration Qualitative Methods Group. <http://cqrmg.cochrane.org/supplemental-handbook-guidance>
- Pace, R., Pluye, P., Bartlett, G., Macaulay, A. C., Salsberg, J., Jagosh, J., & Sellar, R. (2012). Testing the reliability and efficiency of the pilot mixed methods appraisal tool (MMAT) for systematic mixed studies review. *International Journal of Nursing Studies*, 49(1), 47–53. <https://doi.org/10.1016/j.ijnurstu.2011.07.002>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- Peters, A. M. (1977). Language learning strategies: Does the whole equal the sum of the parts? *Language*, 53(3), 560–573. <https://doi.org/10.2307/413177>
- Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide*. Blackwell Pub.
- Pienemann, M., Keßler, J.-U., & Roos, E. (Eds.). (2006). *Englischerwerb in der Grundschule*. UTB.
- Pluye, P., Gagnon, M. P., Griffiths, F., & Johnson-Lafleur, J. (2009). A scoring system for appraising mixed methods research, and concomitantly appraising qualitative, quantitative and mixed methods primary studies in mixed studies reviews. *International Journal of Nursing Studies*, 46(4), 529–546. <https://doi.org/10.1016/j.ijnurstu.2009.01.009>
- Polio, C., & Yoon, H.-J. (2020). Exploring multi-word combinations as measures of linguistic accuracy in second language writing. In B. Le Bruyn & M. Paquot (Eds.), *Learner corpus research meets second language acquisition* (1st ed., pp. 96–121). CUP.
- Richter, K. W. (2021). *Educational outcomes in multilingual CLIL school settings: A systematic review*. Master dissertation, University of Oxford, Oxford, UK. <https://ora.ox.ac.uk/objects/uuid:52221dda-7655-4771-ad7b-66f8c3ee23ca>
- Rixon, S. (2013). *British council survey of policy and practice in primary English language teaching worldwide*. British Council.
- Roehr-Brackin, K. (2014). Explicit knowledge and processes from a usage-based perspective: The developmental trajectory of an instructed L2 learner. *Language Learning*, 64(4), 771–808. <https://doi.org/10.1111/lang.12081>
- Sambanis, M. (2007). *Sprache aus handeln: Englisch und Französisch in der Grundschule*. Verlag Empirische Pädagogik.
- Siyanova-Chanturia, A. (2017). Researching the teaching and learning of multi-word expressions. *Language Teaching Research*, 21(3), 289–297. <https://doi.org/10.1177/1362168817706842>
- Skarabela, B., Ota, M., O'Connor, R., & Arnon, I. (2021). 'Clap your hands' or 'take your hands'? One-year-olds distinguish between frequent and infrequent multiword phrases. *Cognition*, 211, 104612. <https://doi.org/10.1016/j.cognition.2021.104612>
- Slavin, R. E. (1986). Best-evidence synthesis: An alternative to meta-analytic and traditional reviews. *Educational Researcher*, 15(9), 5–11. <https://doi.org/10.3102/0013189X015009005>
- Smith, S. A., & Murphy, V. A. (2015). Measuring productive elements of multi-word phrase vocabulary knowledge among children with English as an additional or only language. *Reading and Writing*, 28(3), 347–369. <https://doi.org/10.1007/s11145-014-9527-y>
- Theakston, A., & Lieven, E. (2017). Multiunit sequences in first language acquisition. *Topics in Cognitive Science*, 9(3), 588–603. <https://doi.org/10.1111/tops.12268>
- Thomas, J., Harden, A., Oakley, A., Oliver, S., Sutcliffe, K., Rees, R., Brunton, G., & Kavanagh, J. (2004). Integrating qualitative research with trials in systematic reviews. *BMJ*, 328(7446), 1010–1012. <https://doi.org/10.1136/bmj.328.7446.1010>
- Thomson, H. (2020). *Developing fluency with multi-word expressions*. Doctoral dissertation, Victoria University of Wellington, Wellington, New Zealand. <https://researcharchive.vuw.ac.nz/handle/10063/9373>
- Tode, T. (2003). From unanalyzed chunks to rules: The learning of the English copula be by beginning Japanese learners of English. *IRAL – International Review of Applied Linguistics in Language Teaching*, 41(1), 23–53. <https://doi.org/10.1515/iral.2003.002>

- Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press.
- Tyler, A. (2010). Usage-based approaches to language and their applications to second language learning. *Annual Review of Applied Linguistics*, 30, 270–291. <https://doi.org/10.1017/S0267190510000140>
- Werlen, E. (2007). *Schlussbericht der Wissenschaftlichen Begleitung WiBe der Pilotphase Fremdsprache in der Grundschule*. Kultusministerium Baden-Württemberg. https://www.km-bw.de/site/pbs-bw/get/documents/KULTUS.Dachmandant/KULTUS/kultusportal-bw/zzz_pdf/Internentfassung.pdf
- Willis, S., Neil, R., Mellick, M. C., & Wasley, D. (2019). The relationship between occupational demands and well-being of performing artists: A systematic review. *Frontiers in Psychology*, 10, 393. <https://doi.org/10.3389/fpsyg.2019.00393>
- Wood, D. (2009). Effects of focused instruction of formulaic sequences on fluent expression in second language narratives: A case study. *Canadian Journal of Applied Linguistics*, 12(1), 39–57.
- Wray, A. (2002). *Formulaic language and the lexicon*. CUP.
- Wray, A. (2013). Formulaic language. *Language Teaching*, 46(3), 316–334.
- Wulff, S. (2018). Acquisition of formulaic language from a usage-based perspective. In A. Siyanova-Chanturia & A. Pellicer-Sánchez (Eds.), *Understanding formulaic language* (pp. 19–38). Routledge.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Schulz, J., Hamilton, C., Wonnacott, E., & Murphy, V. (2023). The impact of multi-word units in early foreign language learning and teaching contexts: A systematic review. *Review of Education*, 11, e3413. <https://doi.org/10.1002/rev3.3413>