The Language Learning Journal

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rllj20

The effect of semantic and physical similarity on vocabulary learning

Nguyen Thien An Bach & Samuel Barclay

To cite this article: Nguyen Thien An Bach & Samuel Barclay (27 Nov 2023): The effect of semantic and physical similarity on vocabulary learning, The Language Learning Journal, DOI: 10.1080/09571736.2023.2286239

To link to this article: https://doi.org/10.1080/09571736.2023.2286239

THE LANGUAGE LEARNING JOURNAL
https://doi.org/10.1080/09571736.2023.2286239

The effect of semantic and physical similarity on vocabulary learning

Nguyen Thien An Bach [a] and Samuel Barclay [b]

[a]Faculty of English Linguistics and Literature, University of Social Sciences and Humanities - Vietnam National University, Ho Chi Minh, Vietnam; [b]Nottingham Institute of Languages and Intercultural Communication, Nottingham Trent University, Nottingham, UK

ABSTRACT

Choosing which words to teach is a key consideration for language teachers and materials writers. Some studies have shown that teaching words in semantically related clusters can make learning more difficult. However, others argue it is the physical similarity of the referents of words that causes confusion. Importantly, studies have employed different methodologies, which may have contributed to the contradictory results. The present study addressed this contradiction by examining the target words from with the methodology of to determine the extent to which semantic relatedness and physical relatedness contribute to the learning burden of L2 vocabulary. Results showed that (1) semantic and physical conditions increased burden relative to an unrelated condition, (2) the greatest burden was found when words were both semantically and physical related, and (3) there was no difference in burden between semantically and physically related words. Pedagogical implications are discussed.

ARTICLE HISTORY
Received 9 June 2023
Accepted 16 November 2023

KEYWORDS
Learning burden; semantic relatedness; physical relatedness; vocabulary learning

Introduction

Vocabulary knowledge is a key part of L2 proficiency (Schmitt and Schmitt 2020), making vocabulary learning a crucial component of any teaching programme (Nation 2013). In many learning contexts textbooks are an important source of vocabulary, shaping the lexical items that learners will encounter and be required to learn. Teaching materials typically present multiple target items in a unit, necessitating the target words be grouped in some manner. It is common for textbooks to group target vocabulary in semantic sets (Nation and Webb 2011) such as colours, animals, or days of the week. Yet empirical research has shown that such sets can increase the learning burden - the amount of effort needed for learning - of L2 vocabulary (Tinkham 1993; Waring 1997) and this practice is generally discouraged by researchers (e.g. Nation 2013). The argument goes that by teaching novel lexical items that share key semantic properties, learners may acquire the different forms, but find it challenging to map those forms to the relevant meanings. The semantic similarity causes interference (Baddeley 1997). However, recent research has argued that the true source of interference is physical rather than semantic relatedness. Physical similarity refers to words whose referents look similar. For instance, Ishii (2015) showed that learning a set of long and thin objects (e.g. *pencil*, *fishing pole*, *chopsticks*, *straw*, *rope, shoelaces*) was more burdensome than a set of kitchen utensils (e.g. pan, knife, cutting board, fork, strainer, ladle).

Crucially, studies in this area have used two methodologies (Nation and Coxhead 2022). The first, a trials-to-criterion design, measures how many repetitions are needed for a set of words

CONTACT Nguyen Thien An Bach anbach@hcmussh.edu.vn

to be learned. The second, a gains-based design, controls the frequency/duration of exposure and measures the amount of learning that occurs (see Barclay and Pellicer-Sánchez 2021 for further discussion). Importantly, investigations that have found an effect for semantic similarity have generally employed a trials-to-criterion design (e.g. Tinkham 1993; Waring 1997), whereas research that has found an effect for physical similarity has used a gains-based design (e.g. Ishii 2015). Thus, the source of disagreement between findings to date may actually stem from differences in research design and thus the true source of interference, semantic or physical, is currently unclear.

Therefore, there is a need to consider the effect of visual similarity using a trials-to-criterion metric to bridge these two research traditions. This paper seeks to address this gap by presenting a study conducted with Vietnamese L1 students who learned target English L2 vocabulary grouped in a variety of conditions, including semantically and physically related sets, using a trials-to-criterion design.

Background

Learning burden

Learning burden relates to the amount of effort required to learn a word (Webb and Nation 2017) and is influenced by linguistic, contextual, and learner-related variables (see Barclay and Schmitt 2019; Higa 1965; Peters 2020), with some factors making the learning burden heavier and others lightening the learning load. One factor that has received considerable research attention is the grouping of target items. This factor is directly relevant to teaching and learning, where target items are rarely presented in isolation; for instance, textbooks often present a list of specific target items at the beginning of a unit (e.g. Lebeau and Rees 2008; Schmitt et al. 2011) or on review pages at the end of a unit (e.g. de Chazal and Rogers 2013). Studies have considered the effect of the order in which items are presented (Lado et al. 1967) as well as the relationship between the target items (Ishii 2015; Tinkham 1993; Waring 1997), the focus of the current paper.

Interference

One concern when target items are presented together is that the learning of one item may impact the learning of other items in the learning set. Interference Theory (see Crowder 1976) holds that learning burden relates to the level of similarity between target items, with increased similarity making learning more burdensome. Research has shown that words that are similar in form (Laufer 1989) and/or meaning (Tinkham 1997) are harder to learn when they are presented at the same time. This effect is thought to stem from interference (Nation 2013), which occurs when learners inadvertently map the form of one word to the meaning of another. For example, if *right* and *left* are taught together, while learners may know that these forms refer to directions, they may be unsure which form refers to which direction. Interference can also arise from near semantic and formal relationships such as synonymy (e.g. *progress* and *advance*) (Webb and Nation 2017), synformy (similar forms e.g. *price* and *prize*) (Laufer 1997), and semantic sets (e.g. numbers, days of the week) (Tinkham 1993; Waring 1997). It has been claimed that interference can increase the learning burden of novel vocabulary by as much as 100% (Nation 2000), so researchers recommend presenting related words independently (Webb and Nation 2017). Such practice might diverge from the intuition of some language teachers who often consider grouping items in semantic sets to be supportive for vocabulary learning (Ishii 2015). Crucially, however, there is currently some debate over the source of interference when items are presented in semantic clusters (discussed in the next section). More research is therefore needed to clarify this area, aiding the development of evidence-based practice and informing the design of maximally effective teaching materials.

Presenting words in semantic clusters

Sets of words that share a common superordinate concept and have certain syntactically and semantically shared properties have been variously described as *lexical sets* (Gairns and Redman 1986), *semantic sets* (Waring 1997), and *semantic clusters* (Tinkham 1993). The prevalence of such sets has been observed in a number of English course books (Erten and Tekin 2008; Waring 1997) as it is convenient for materials developers (Folse 2004; Tinkham 1993) and often thought to facilitate vocabulary learning (Finkbeiner and Nicol 2003).

Although some studies have shown that clustering semantically related items facilitates learning (Hashemi and Gowdasiaei 2005; Hoshino 2010), the majority of research in this area has found such clustering to increase the learning burden of L2 vocabulary (Bak 2012; Bolger and Zapata 2011; Erten and Tekin 2008; Finkbeiner and Nicol 2003; Jang 2014; Tinkham 1993; Waring 1997). For instance, in a seminal study, Tinkham (1993) used a trials-to-criterion methodology to compare the learning of paired associates (English words and nonwords) presented in semantically related and unrelated conditions. Results showed that more trials were needed in the semantically related condition and participants also reported that learning the semantically related items was more difficult. This study was replicated in a Japanese context by Waring (1997) , who also found that semantic clustering hindered vocabulary acquisition. In a later study, Tinkham (1997) also found that semantic clustering caused more burden than thematic clustering (i.e. grouping items by theme: *frog, pond, hop*). He thus suggested that materials writers adopt a thematic approach to the selection of target items.

From a theoretical perspective, Hunt and Elliot (1980) argue that the learning burden of a target item relates to its distinctiveness, with more distinctive items easier to learn. Thus, words presented in semantic clusters are harder to learn because the constituent items are harder to distinguish from one another, with learners less able to identify salient features of individual items.

A new perspective on cross-association

Ishii (2015) brought a new perspective to this area by recognising that many of the target words in previous studies had physically similar, as well as semantically similar, referents. For instance, fruits such as *peach, nectarine,* and *apple* as well as items of clothing such as *shirt*, *jacket*, and *sweater*, target items from Tinkham's 1993 study, are both physically similar and semantically related. There is reason to believe that such physical similarity may be particularly impactful on vocabulary learning, with research showing that using pictures to illustrate word meaning positively affects acquisition (Bisson et al. 2015) and theories of human memory ascribing a key role to visual information (e.g. Paivio 1969).

In a series of studies with Japanese L1 learners of English, Ishii investigated the extent to which physical similarity superseded the impact of semantic relatedness, determining whether the effect for semantic similarity reported in previous studies may have actually stemmed from the shared physical characteristics of the target referents. In her first study, Ishii (2015) compared three conditions: semantically unrelated but physically related, semantically related but physically unrelated, and both semantically and physically unrelated. Using a gains-based design, she found that when semantic relatedness was controlled, physical relatedness led to significantly less learning (meaning recall) than the other two conditions. For example, the participants learned fewer semantically unrelated round objects (e.g. *globe; watermelon; ball*) than semantically related (e.g. *monkey; snake; giraffe)* or unrelated items (e.g. *forest; cup; scissors)*. Additionally, she found that when physical relatedness was controlled, semantic relatedness did not impact learning.

In her 2017 study, Ishii investigated the relative burden of learning semantically related and unrelated abstract nouns, which by nature are not physically related. Again using a gains-based approach, she found no statistically significant difference between the two target-item groups, showing that semantic relatedness did not impact the learning of abstract words.

In her most recent study (2022), Ishii compared the learning of physically and semantically related sets with physically unrelated but semantically related sets. In a gains-based approach, using a L1-pseudoword paired-associate learning task, she found that physical relatedness caused more cases of confusion, leading her to conclude that 'it is hard to deny the possible involvement of visual imagery in the efficacy of learning semantically related words' (Ishii 2022: 189).

Importantly, however, Ishii employed a different methodology to the studies of Tinkham (1993) and Waring (1997), which may have impacted the results. Tinkham (1993) used a trials-to-criterion metric whereas Ishii used a gains-based metric. A trials-to-criterion metric measures how many repetitions are needed for learning; it considers the learning process. A gains-based approach uses a pre- and post-test design, or a post-test if nonwords are used, measuring the amount of learning that occurs and using this data to infer the relative difficulty of the learning process. Recent studies (Barclay and Pellicer-Sánchez 2021) have argued that trials-to-criterion approaches are preferable when measuring learning burden because under a gains-based approach, it is not possible to determine the relative difficulty of each learned item; we only know whether an item has been learned or not. Trials-to-criterion approaches can offer a more nuanced picture of learning burden. The study described in this paper used a trials-to-criterion approach to consider the effect of semantic and physical relatedness on the learning of novel L2 lexical items. Ishii (2022: 188) recognised the potential contribution of such a study, stating that 'more detailed tracking of the memorization process, such as keeping a log of time spent on each item or the number of times each item is reviewed, might reveal more about the actual extent of confusion [caused by the co-presentation of physically related items]'.

This study

Given the general lack of consensus and the methodological differences between studies to date, it is currently unclear whether the source of interference lies in semantic and/or physical similarity. Thus, there is a need for a study to conceptually replicate (see Porte and Richards 2012 for a discussion of conceptual replication) Ishii (2015, 2022) using Tinkham's (1993) methodology to determine whether the apparent contradiction between these studies stems from paradigmatic or methodological differences. The current study sought to meet this gap by investigating the learning of L1-pseudoword paired associates in a Vietnamese higher education context. In the first experiment, we replicated Ishii (2022), seeking to validate the finding that when words are semantically related, physical relatedness increases the learning burden of novel lexical items. In the second experiment, we investigated Ishii's (2015) finding that physically related, but semantically unrelated, items are harder to learn than physically unrelated, but semantically related lexical items. The study considered the following two research questions:

1. To what extent do semantic relatedness and physical relatedness impact the learning burden of L2 lexical items?
2. For semantically grouped words, to what extent does physical relatedness impact the learning burden of L2 lexical items?

Methodology

The participants

L1 Vietnamese English learners took part in this study (age: $M = 23.42$ $SD = 6.13$). Participants ($N = 72$) were students at either universities or private language institutions in Vietnam at the time of data collection, and were recruited using a snowball sampling method. Participants had recently taken the IELTS test, and these scores were used to ensure a homogeneous sample in terms of proficiency. The IELTS scores indicated that the learners were at the B2 level of the Common European

Framework of Reference for Languages (Council of Europe 2020). This study adopted a within-subjects design, but to control for the test form, learners were split into three groups and were presented with the target items in a different order (see Figure 1).

Target items

The target items (L1 equivalents and nonwords) were taken from Ishii (2015), but as her participants were L1 Japanese learners of English and the current study was conducted with L1 Vietnamese learners, equivalents were translated into Vietnamese. This was performed by an L1 user of Vietnamese highly proficient in English and all translations were checked by a second translator. Additionally, one equivalent 'Japanese radish' was changed to a more contextually appropriate equivalent as the translators judged this item likely to be unknown to many participants. Target items are given in Appendix 1.

Instruments

Experiment 1

Experiment 1 required participants to study nine word pairs to criterion. Three items were semantically clustered but physically unrelated, three items were semantically clustered and physically related, and three items were neither semantically nor physically related. The items were presented sequentially. Three forms of the test were created to allow for the possibility that some nonwords were easier to learn than others. For instance, the nonwords paired with the unrelated words on one form were assigned to semantically related words on another test form and physically related words on the final form. Table 1 shows the word-pairs used on Form A (see Appendix 2 for Forms B and C).

Experiment 2

In Experiment 2, participants learned three sets of six paired associates, one set with semantically and physically unrelated items, one with semantically but not physically related items, and one with physically but not semantically related items. Three forms of the test were created to control for any unforeseen differences in the difficulty of the target nonwords. Thus, the specific non-word L1 equivalent combinations varied by form, with nonwords paired with unrelated words on one form assigned to, for example, the semantically related set on another test form. Table 2 illustrates the construction of the three sets for Form A (see Appendix 2 for Forms B and C). In addition, to avoid a potential order effect, the order in which the sets were presented varied, with a third of participants encountering the unrelated set first, the semantically related set second, and the physically related set last. Another third encountered the physically related set first, the unrelated set second, and the

Target word set presentation order by group

Group A (n = 24)	Group B (n = 24)	Group C (n = 24)
Unrelated set	Semantically related set	Physically + semantically related set
Semantically related set	Physically + semantically related set	Unrelated set
Physically + semantically related set	Unrelated set	Semantically related set

Figure 1. Target word set presentation order by group.

Table 1. Experiment 1 – target word pairs (Form A).

Unrelated	Semantically related	Physically and semantically related
cái muỗng (spoon) – parn	con heo (pig) – forl	áo khoác (jacket) – cale
quả anh đào (cherry) – hoff	con gà (chicken) – cudy	áo sơ mi (shirt) – broof
ngọn núi (mountain) – sarny	hươu cao cổ (giraffe) – greap	áo choàng (coat) – damble

semantically related set last. The final third were exposed to the semantically related set first, the physically related set second, and the unrelated set last.

Procedure

Following Tinkham (1993), the participants first listened to recorded instructions and signed a consent form. After this, they completed a practice trials-to-criterion test and were able to ask the researcher any questions. Once all questions had been answered, the session began with Experiment 1 and then immediately continued to Experiment 2. The data were collected online, via a popular web conferencing software. Data collection was conducted one learner at a time. Participants had their web cameras on, and the image included their hands to ensure they were not taking notes. Prior to data collection, we confirmed that participants were alone in a quiet location, had a stable internet connection, and could clearly hear the target items.

In the first experiment, the participants learned a set of nine pairs (i.e. three unrelated pairs, three semantically related pairs, and three semantically and physically related pairs), with the conditions presented sequentially. Participants were exposed to the spoken forms of the nonwords and the L1 equivalents were also read out. They heard the L1 equivalent of each target item first followed by the relevant nonword two seconds later. Both the L1 equivalent and the nonword were read by an L1 user of Vietnamese proficient in English. After the learning phase, participants attempted to recall the target items, producing the correct nonword in response to an L1 equivalent stimulus within a 3-second period. Thus, we followed the methodology of Tinkham (1993) conducting a timed, spoken form recall task. The researcher listened and recorded the accuracy of production. For each nonword recalled accurately, the researcher silently put a tick on a mark sheet. Accrued items were retained in subsequent trials, so that each trial contained the same number of target items. There was a two-second interval between word pairs. After they had worked through all the pairs, they repeated the learning-testing routine until all items were successfully recalled. There was a five-second pause between trials. The test concluded when the participants had learned all the pairs in the set. For all pairs, the individual items were reordered from one trial to the next to avoid a sequence effect. This ordering was carefully considered so that each pair could appear in various positions within the set and the last pair of one trial did not become the first pair of the following trial, preventing a recency effect. The number of trials needed to reach criterion (i.e. successful recall of all target spoken forms in a set) was recorded and used as a metric for learning burden.

The participants then completed the second experiment. Participants were presented with three sets of items organised by condition (i.e. semantically and physically unrelated; physically unrelated but semantically related set; and physically related but semantically unrelated) and completed the

Table 2. Experiment 2 – target word pairs (Form A).

Unrelated set	Semantically related set	Physically related set
cục gôm (eraser) – nein	củ cải (radish) – count	quả địa cầu (globe) – boak
con sư tử (lion) – marby	dưa leo (cucumber) – heam	dưa hấu (watermelon) – dar
đàn dương cầm (piano) – cank	rau dền (spinach) – ligs	quả bóng (ball) – groop
đôi vớ (socks) – rarty	đậu bắp (okra) – paff	ngọc trai (pearl) – modds
trái cam (orange) – mamp	cà chua (tomato) – stace	viên kẹo (candy) – stalt
ấm nước (kettle) – stass	cà tím (eggplant) – woak	viên bi (marble) – tice

same trials-to-criterion process as Experiment 1. Target pairs were presented in the same manner as Experiment 1. When all the constituent items in a condition were accurately recalled, the number of trials needed to reach that criterion was noted and the participant moved to the next condition. Experiment 2 concluded once all three conditions had been completed.

Data analysis

The data of both experiments were analysed in SPSS using mixed ANOVAs with the outcome variable the number of trials needed to reach criterion. Mixed analysis of variance combines both between-subjects (i.e. test form) and within-subjects (i.e. condition) variables into one analysis (Field 2018). The data were log transformed prior to analysis to avoid issues with equality of covariance and homogeneity of variance (see Field 2018). All assumptions associated with Mixed ANOVAs (see Field 2018) were met. Post-hoc Tukey tests were run to understand potential differences between the three test forms and the three experimental conditions.

Results

Experiment 1

Descriptive statistics for Experiment 1 are presented in Table 3. The data show the average number of trials needed for items to be encoded in the three conditions. In general, it took an average of between four and five trials to successfully encode the target items. Inferential statistics showed that there was no statistically significant interaction between the form and condition (Wilks' Lambda = .88, F (4, 136) = 2.27, p = .07. partial eta squared = .06), nor was there any effect for test form (F (2, 69) = .10, p = .90, partial eta squared < .01), indicating that test form did not impact the results. However, there were important differences in the number of trials needed to learn items in the various conditions, Wilks' Lambda = .60, F (2, 68) = 22.97, p < .001, partial eta squared = .40. Pairwise comparisons revealed that unrelated words required significantly fewer trials than semantically related (p < .03, d = .36) and semantically + physically related words (p < .001, d = .68). Importantly, a significant difference was also found between semantically related and semantically + physically related items (p = .001, d = .42), indicating that physical relatedness significantly contributed to the learning burden of semantically related L2 vocabulary.

Experiment 2

The descriptive statistics associated with Experiment 2 are presented in Table 4. Again, the average number of trials needed to reach criterion is presented by condition. In general, items needed between four and seven trials to reach criterion. Inferential statistics showed that there was no statistically significant interaction between test form and condition (Wilks' Lambda = .90, F (4, 136) = 1.76, p = .14, partial eta squared = .05) nor any effect for test form (F (2, 69) = 1.46, p = .24, partial eta squared = .04). However, a statistically significant effect for instructional condition was found Wilks' Lambda = .55, F (2, 68) = 28.29, p < .001, partial eta squared = .45). Pairwise comparisons indicated that the unrelated set required significantly fewer trials than both the semantically (p < .001, d = .71) and physically (p < .001, d = .53) related items. Importantly, there was no statistically significant

Table 3. Trials to criterion for experiment 1 by condition.

Condition	Minimum	Maximum	Mean	*SD*	*SEM*
Unrelated	1	9	4.17	2.06	.24
Semantic	1	13	4.69	1.88	.22
Physical + Semantic	1	13	5.90	2.61	.31

Note. N = 72.

Table 4. Trials to criterion for experiment 2 by condition.

Condition	Minimum	Maximum	Mean	*SD*	*SEM*
Unrelated	2	13	4.93	2.49	.29
Semantic	1	15	6.76	3.06	.36
Physical	1	15	6.33	3.02	.36

difference in the number of trials needed in the semantic and physical relatedness condition ($p = .27$), indicating that physical relatedness did not pose more burden than visual relatedness.

Discussion

Across two experiments, this study considered the discrete contributions of semantic and physical relatedness to the learning burden of L2 vocabulary. It also investigated the additional contribution of physical similarity to the learning burden of semantically related words. Here the results are discussed according to the research questions this paper sought to answer.

RQ1: to what extent do semantic relatedness and physical relatedness impact learning burden?

The results showed that items in the unrelated condition were learned with fewer trials than those in the semantically related condition (Experiments 1 and 2), indicating that presenting target items in semantically related sets led to an increased learning burden. This finding is in line with previous research that has shown such presentation to be detrimental to learning (e.g. Tinkham 1993; Waring 1997). Interestingly, however, the data showed that the average increase in difficulty was not as dramatic as some scholars have previously estimated (e.g. Nation 2000), with semantic relatedness causing a 12.5% increase in learning burden in Experiment 1 and a 37.1% increase in Experiment 2. The greater increase in burden in Experiment 2 was likely caused by more target items being employed per condition (i.e. six) compared to Experiment 1 (i.e. three). Importantly, this finding contradicts the results of Ishii (2015) who found no effect for semantic relatedness when physical similarity was controlled. Thus, it seems that the methodological differences between previous studies may, to some extent, explain the contradictory results reported to date.

The findings (Experiment 2) also showed that unrelated words were easier to learn than items grouped according to physical relatedness, with physical relatedness causing a 28.4% increase in burden compared to the unrelated condition. Thus, with regard to physical relatedness, the current study supports the findings of Ishii (2015, 2017, 2022) in demonstrating that presenting items in sets according to the physical characteristics of their referents does, demonstrably, make learning more burdensome.

RQ2: for semantically grouped words, to what extent does physical relatedness impact learning burden?

The results (Experiment 1) indicated that when the target items were semantically related, additional physical relatedness increased the learning burden. For instance, words grouped in unrelated sets required fewer trials than items grouped in semantically and physically related clusters, with physical and semantic relatedness leading to a 41.5% increase in burden compared to the unrelated set. Importantly, the data (Experiment 1) showed that items presented in the semantically related condition posed less burden than items presented in the semantically and physically related condition. Physically and semantically related items were 25.8% more burdensome than items in the semantic condition. Thus, it seems that physical relatedness does make a unique contribution to the learning burden of target items, beyond that caused by semantic relatedness.

Crucially, however, the size of such a contribution does not seem to differ greatly from that of semantic relatedness, meaning that physical relatedness is not more burdensome than semantic relatedness. There was no statistically significant difference (Experiment 2) between the number of trials needed in the semantically related condition and the physically related condition. Thus, it seems that physical similarity does cause burden, but the extent of that burden is no greater than that caused by semantic similarly. This finding contradicts Ishii's study of 2015, which reported that physically related items were more burdensome than semantically related items. This difference likely stems from the contrasting methodologies employed, with Ishii adopting a gains-based approach and the current study using a trials-to-criterion approach. Importantly, it has recently been argued (Barclay and Pellicer-Sánchez 2021) that trials-to-criterion metrics better capture the nature of learning burden as they measure the learning process directly rather than using gains to infer the relative difficulty of learning. Thus, the findings suggest that the role of physical similarity might not be as significant as has previously been thought.

Limitations

There are some limitations associated with this study. First, in Experiment 1 there were only three pairs per condition which may have impacted the amount of burden experienced. This design was based on Tinkham (1993) and so was justified given the replicatory focus of the current study. However, as we would expect more burden in the related conditions with an increase in target items, which may expose differences between the various conditions, future studies might employ more target items. Secondly, to control for prior knowledge of the target items, we used nonwords. This choice was necessary given that Ishii's (2015) original study also used nonwords; however, it may be that the use of nonwords impacted the findings. This is particularly the case because when selecting nonwords, we are able to control several intralexical variables such as word length. Doing so aids the reliability of the study as it controls for potential confounds, but it may also impact results as the forms presented to learners are somewhat similar. This lack of formal salience may increase the burden of the target items (Laufer 1989). Research is needed in this area to inform the design of future studies. Third, the target items were all nouns that related to concrete objects. Such items are easier to learn than less visually salient abstract nouns and other parts of speech (Ellis and Beaton 1993). Therefore, the findings may not generalise to other types of target item and further research is needed. Fourth, following Ishii (2015) we did not conduct a delayed post-test; it is therefore not possible to determine the long-term impact of presenting items in semantically or physically related sets. Thus, research employing delayed post-tests is needed to clarify this area. Additionally, while the manner of intentional learning used in this study reflected the methodology of Tinkham (1993), it likely differed from typical learner behaviour, which often employs the written form via word lists and flashcards. It is possible that the presence of the written form may impact the learning burden of the target items; thus, the findings might not directly transfer to the classroom. Finally, the findings may have been impacted by interlexical and learner variables, meaning that the results may not generalise to non-Vietnamese L1 learners and students with different proficiency levels.

Implications and conclusion

This paper has presented a study that investigated the effect of physical and semantic similarity on the learning burden of L2 vocabulary. It has shown that both semantic and physical clustering increase the learning burden of L2 lexical items, with items presented in a combined condition (i.e. semantic and physical clustering) found to be even more problematic for learners. Importantly, no difference between the burden posed by semantic and physically related items was found. This

conceptual replication employed the target items from Ishii (2015, 2022) and the methodology of Tinkham (1993) in an attempt to clarify some areas of conflict in findings to date. The results of this paper support both Tinkham (1993) and Ishii (2015) in that semantic and physical relatedness were shown to impact burden. However, the findings contradict Ishii (2015) as (1) no difference was found between the burden posed by the two conditions and (2) semantic relatedness was found to impact burden.

These findings have several important implications for language teaching. First, as we have shown that grouping physically and semantically related items can increase learning burden, teachers and materials writers should avoid presenting words which have visually similar referents, or whose meanings belong to the same superordinate category, at the same time. They should be especially wary of teaching semantically *and* physically related words together, as the results of Experiment 2 showed this to be particularly taxing for the learners. The findings also indicated that grouping unrelated items together produced the lightest burden. Following such an approach can be a challenge for materials writers; however, one strategy might be to focus on word frequency (see Nation 2013 for a discussion on the importance of word frequency in lexical selection), selecting items based on how commonly they occur in language use, rather than whether or not their meanings can be included in a specific topic or category. Alternatively, writers could adopt Tinkham's (1997) suggestion, grouping words around a story or theme.

Methodologically, this study has also demonstrated that a trials-to-criterion approach can be used to measure learning burden, allowing us to consider burden as a function of the learning process. Researchers investigating burden might consider using such a design over gains-based approaches. Recent research has demonstrated how this approach can be operationalised in an ecologically valid manner, with minimal disturbance to natural learning behaviour (see Barclay and Pellicer-Sánchez 2021).

Overall, this study has demonstrated that careful consideration needs to be given to the grouping of target items to ensure that they are learned with maximum efficiency. We have demonstrated that two variables, semantic and physical relatedness, impact the learning burden of L2 vocabulary and so need to be controlled in the design of language teaching materials. Such research is important if we are to develop evidence-based practice and avoid introducing unnecessary burden into the language classroom.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Nguyen Thien An Bach http://orcid.org/0000-0002-9245-3298
Samuel Barclay http://orcid.org/0000-0001-6112-382X

References

Baddeley, A.D. 1997. *Human Memory: Theory and Practice*. Revised ed. London: Psychology Press.

Bak, S. 2012. The effects of semantic relatedness on EFL vocabulary recall and retention. *English Teaching* 67, no. 3: 281–305. doi:10.15858/engtea.67.3.201209.281

Barclay, S. and A. Pellicer-Sánchez. 2021. Exploring the learning burden and decay of foreign language vocabulary knowledge: the effect of part of speech and word length. *ITL-International Journal of Applied Linguistics* 172, no. 2: 259–289. doi:10.1075/itl.20011.bar

Barclay, S. and N. Schmitt. 2019. Current perspectives on vocabulary teaching and learning. In *Springer Handbook of English Language Teaching*, ed. X. Gao, 799–819. Cham: Springer.

Bisson, M.J., W.J. van Heuven, K. Conklin and R.J. Tunney. 2015. The role of verbal and pictorial information in multimodal incidental acquisition of foreign language vocabulary. *Quarterly Journal of Experimental Psychology* 68, no. 7: 1306–1326. doi:10.1080/17470218.2014.979211

Bolger, P. and G. Zapata. 2011. Semantic categories and context in L2 vocabulary learning. *Language Learning* 61, no. 2: 614–646. doi:10.1111/j.1467-9922.2010.00624.x

Council of Europe. 2020. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment - Companion Volume*. Strasbourg: Council of Europe Publishing.

Crowder, R.G. 1976. *Principles of Learning and Memory*. Mahwah, NJ: Lawrence Erlbaum.

de Chazal, E. and L. Rogers. 2013. *Oxford EAP Intermediate: A Course in English for Academic Purposes*. Oxford, UK: Oxford University Press.

Ellis, N. and A. Beaton. 1993. Factors affecting the learning of foreign language vocabulary: imagery keyword mediators and phonological short-term memory. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology* 46A, no. 3: 533–558. doi:10.1080/14640749308401062

Erten, I.H. and M. Tekin. 2008. Effects on vocabulary acquisition of presenting new words in semantic sets versus semantically unrelated sets. *System* 36, no. 3: 407–422. doi:10.1016/j.system.2008.02.005

Field, A. 2018. *Discovering Statistics using IBM SPSS Statistics*. 5th ed. London, UK: SAGE.

Finkbeiner, M. and J. Nicol. 2003. Semantic category effects in second language word learning. *Applied Psycholinguistics* 24, no. 3: 369–383. doi:10.1017/S0142716403000195

Folse, K.S. 2004. Myths about teaching and learning second language vocabulary: what recent research says. *TESL Reporter* 37, no. 2: 1–13.

Gairns, R. and S. Redman. 1986. *Working with Words: A Guide to Teaching and Learning Vocabulary*. Cambridge, UK: Cambridge University Press.

Hashemi, M.R. and F. Gowdasiaei. 2005. An attribute-treatment interaction study: lexical-set versus semantically-unrelated vocabulary instruction. *RELC Journal* 36, no. 3: 341–361. doi:10.1177/0033688205060054

Higa, M. 1965. The psycholinguistic concept of 'difficulty' and the teaching of foreign language vocabulary. *Language Learning* 15, no. 3–4: 167–179. doi:10.1111/j.1467-1770.1965.tb00799.x

Hoshino, Y. 2010. The categorical facilitation effects on L2 vocabulary learning in a classroom setting. *RELC Journal* 41, no. 3: 301–312. doi:10.1177/0033688210380558

Hunt, R.R. and J.M. Elliot. 1980. The role of nonsemantic information in memory: orthographic distinctiveness effects on retention. *Journal of Experimental Psychology: General* 109, no. 1: 49–74. doi:10.1037/0096-3445.109.1.49

Ishii, T. 2015. Semantic connection or visual connection: investigating the true source of confusion. *Language Teaching Research* 19, no. 6: 712–722. doi:10.1177/1362168814559799

Ishii, T. 2017. The impact of semantic clustering on the learning of abstract words. *Vocabulary Learning and Instruction* 6, no. 1: 21–31.

Ishii, T. 2022. The role of visual imagery in semantic clustering. In *Vocabulary Theory, Patterning, and Teaching*, ed. P. Szudarski and S. Barclay, 171–187. Bristol, UK: Multilingual Matters.

Jang, H.J. 2014. The effects of semantic clustering on EFL young learners' vocabulary learning. *English Teaching* 69, no. 3: 25–47. doi:10.15858/engtea.69.3.201409.25

Lado, R., B. Baldwin and F. Lobo. 1967. *Massive Vocabulary Expansion in a Foreign Language beyond the Basic Course: The Effects of Stimuli, Timing and Order of Presentation*. Washington, D.C. U.S. Department of Health, Education, and Welfare, Project No. 5-1095.

Laufer, B. 1989. A factor of difficulty in vocabulary learning: deceptive transparency. *AILA Review* 6, no. 1: 10–20.

Laufer, B. 1997. The lexical plight in second language reading: words you don't know, words you think you know, and words you can't guess. In *Second Language Vocabulary Acquisition: A Rationale for Pedagogy*, ed. J. Coady and T. Huckin, 20–34. Cambridge, UK: Cambridge University Press.

Lebeau, I. and G. Rees. 2008. *Elementary Language Leader Coursebook*. London, UK: Pearson Longman.

Nation, I.S.P. 2000. Learning vocabulary in lexical sets: dangers and guidelines. *TESOL Journal* 9, no. 2: 6–10. doi:10.1002/j.1949-3533.2000.tb00239.x

Nation, I.S.P. 2013. *Learning Vocabulary in Another Language*. 2nd ed. Cambridge, UK: Cambridge University Press.

Nation, I.S.P. and A. Coxhead. 2022. Vocabulary learning and teaching. In *Vocabulary Theory, Patterning, and Teaching*, ed. P. Szudarski and S. Barclay, 171–187. Bristol, UK: Multilingual Matters.

Nation, I.S.P. and S. Webb. 2011. *Researching and Analyzing Vocabulary*. Boston, MA: Heinle Cengage Learning.

Paivio, A. 1969. Mental imagery in associative learning and memory. *Psychological Review* 76, no. 3: 241–263. doi:10.1037/h0027272

Peters, E. 2020. Factors affecting the learning of single-word items. In *The Routledge Handbook of Vocabulary Studies*, ed. S. Webb, 125–142. Abingdon, UK: Routledge.

Porte, G. and K. Richards. 2012. Focus article: replication in second language writing research. *Journal of Second Language Writing* 21, no. 3: 284–293. doi:10.1016/j.jslw.2012.05.002

Schmitt, N. and D. Schmitt. 2020. *Vocabulary in Language Teaching*. 2nd ed. Cambridge, UK: Cambridge University Press.

Schmitt, D., N. Schmitt and D. Mann. 2011. *Focus on Vocabulary 1: Bridging Vocabulary*. White Plains, NY: Pearson Education.

Tinkham, T. 1993. The effect of semantic clustering on the learning of second language vocabulary. *System* 21, no. 3: 371–380. doi:10.1016/0346-251X(93)90027-E

Tinkham, T. 1997. The effects of semantic and thematic clustering on the learning of second language vocabulary. *Second Language Research* 13, no. 2: 138–163. doi:10.1191/026765897672376469
Waring, R. 1997. The negative effects of learning words in semantic sets: a replication. *System* 25, no. 2: 261–274. doi:10.1016/S0346-251X(97)00013-4
Webb, S. and I.S.P. Nation. 2017. *How Vocabulary is Learned*. Oxford University Press.

Appendix 1

Table A1. Target L1 equivalents and their meanings in English.

Vietnamese equivalents	Meaning in English
cái muỗng	a spoon
quả anh đào	a cherry
ngọn núi	a mountain
con heo	a pig
con gà	a chicken
hươu cao cổ	a giraffe
áo khoác	a jacket
áo sơ mi	a shirt
áo choàng	a coat
cục gôm	an eraser
củ cải	a radish
quả địa cầu	a globe
con sư tử	a lion
đàn dương cầm	a piano
đôi vớ	socks
trái cam	an orange
ấm nước	a kettle
dưa hấu	a watermelon
quả bóng	a ball
ngọc trai	a pearl
viên kẹo	a candy
viên bi	a marble

Table A2. The target nonwords.

boak	damble	mamp	rarty
broof	dar	marby	sarny
cale	forl	modds	stalt
cank	greap	nein	stass
count	groop	parn	tice
cudy	hoff		

Appendix 2

Table A3. Experiment 1 – target word pairs (Form B).

Unrelated	Semantically related	Physically and semantically related
cái muỗng (spoon) – forl	con heo (pig) – cale	áo khoác (jacket) – parn
quả anh đào (cherry) – cudy	con gà (chicken) – broof	áo sơ mi (shirt) – hoff
ngọn núi (mountain) – greap	hươu cao cổ (giraffe) – damble	áo choàng (coat) – sarny

Table A4. Experiment 1 – target word pairs (Form C).

Unrelated	Semantically related	Physically and semantically related
cái muỗng (spoon) – cale	con heo (pig) – parn	áo khoác (jacket) – forl
quả anh đào (cherry) – broof	con gà (chicken) – hoff	áo sơ mi (shirt) – cudy
ngọn núi (mountain) – damble	hươu cao cổ (giraffe) – sarny	áo choàng (coat) – greap

Table A5. Experiment 2 – target word pairs (Form B).

Unrelated set	Semantically related set	Physically related set
cục gôm (eraser) – count	củ cải (radish) – boak	quả địa cầu (globe) – nein
con sư tử (lion) – heam	dưa leo (cucumber) – dar	dưa hấu (watermelon) – marby
đàn dương cầm (piano) – ligs	rau dền (spinach) – groop	quả bóng (ball) – cank
đôi vớ (socks) – paff	đậu bắp (okra) – modds	ngọc trai (pearl) – rarty
trái cam (orange) – stace	cà chua (tomato) – stalt	viên kẹo (candy) – mamp
ấm nước (kettle) – woak	cà tím (eggplant) – tice	viên bi (marble) – stass

Table A6. Experiment 2 – target word pairs (Form C).

Unrelated set	Semantically related set	Physically related set
cục gôm (eraser) – boak	củ cải (radish) – nein	quả địa cầu (globe) – count
con sư tử (lion) – dar	dưa leo (cucumber) – marby	dưa hấu (watermelon) – heam
đàn dương cầm (piano) – groop	rau dền (spinach) – cank	quả bóng (ball) – ligs
đôi vớ (socks) – modds	đậu bắp (okra) – rarty	ngọc trai (pearl) – paff
trái cam (orange) – stalt	cà chua (tomato) – mamp	viên kẹo (candy) – stace
ấm nước (kettle) – tice	cà tím (eggplant) – stass	viên bi (marble) – woak