Social Sciences & Humanities Open 7 (2023) 100398

Contents lists available at ScienceDirect

Social Sciences & Humanities Open

journal homepage: www.sciencedirect.com/journal/social-sciences-and-humanities-open

Regular Article

Normative data for logographic and lexical Japanese paired associates

Wesley Pyke [a], Johan Lunau [a], Amir-Homayoun Javadi [a,b,*]

[a] School of Psychology, University of Kent, Canterbury, UK
[b] School of Rehabilitation, Tehran University of Medical Sciences, Tehran, Iran

ARTICLE INFO

Keywords:
Pair-word associates
Logographic
Lexical
Abstract stimuli
Recall accuracy
Item difficulty

ABSTRACT

The current study aimed to provide normative paired-associate data for an unexplored logographic language system (Japanese). 240 participants with a mean age of 21.8 ($SD = 3.6$) were randomly allocated to one of three conditions in an independent groups design: logographic (Japanese symbol - English word), lexical (Japanese word - English word), or abstract paired associates (Japanese symbol - Japanese word). Participants underwent three study-test trials in each condition. Recall accuracy was found to increase across trials whilst relative item difficulty (correlations between trials in each condition) remained similar. A simultaneous multiple regression revealed that Japanese symbol stroke count was the only reliable and significant predictor of recall accuracy. Metacognitive measures, when correlated with recall accuracy, showed that participants were generally aware of how well they had learned the paired associates. Based on prior research into memory for pictures and words, we hypothesized that the recall accuracy of participants in the logographic condition would exceed that of those in the lexical condition. This hypothesis was rejected following a post hoc independent *t*-test comparing the two groups. This study provides a thorough normative data set that can be customized by researchers for diverse purposes in addition to novel insights, such as revealing participants' recall accuracy for abstract paired associates for which they had no previous semantic representations.

1. Introduction

Paired associate tasks are a popular and widely recognised method of studying human memory. In a typical paired-associate task, the learner is presented with a list of paired items such as words, images, or numbers to study. Following this study phase, the learner's recall is assessed by means of a memory test that presents only one item of each pair (the stimulus), which is intended to serve as a cue for recall of the second item (the response). The theory underlying paired-associate tasks is that behavioural stimulus-response associations are formed between the item pairs such that the presentation of one item (the stimulus) initiates recall of the other (the response), fostering learning.

Researchers have sought to develop normative data sets for foreign language-English translation equivalents to provide a standardized stimuli pool for further research. In a pioneering study, Nelson and Dunlosky (1994) provided normative data for Swahili-English translation equivalents. Their rationale behind the employment of these languages is clear and intuitive, and it explains why the use of this language pair dominates the literature on memory research (Jang & Nelson, 2005; Kelemen et al., 2007; Richards & Nelson, 2004; Toppino & Cohen, 2009; Van Overschelde & Nelson, 2006). To elaborate, Swahili uses the same standard English alphabet without letter modifications such as accents; English-speaking participants are unlikely to have been exposed to it or to have learned it in academic settings; Swahili-English word pairs neither produce a ceiling nor a floor effect, permitting effective assessment of learning; and Swahili is not a Romance language, minimizing the occurrence of confounding cognates, which are words that have a shared etymological origin. For example, the German word *Tisch* (table), the English words dish, desk, and disk, and the Latin word *discus* are cognates because they have the same parent language - Ancient Greek. However, Nelson and Dunlosky's (1994) study and the use of Swahili-English translation equivalents in general is subject to limitations. Indeed, the frequent use of Swahili and English translation equivalents across multiple experiments creates the dilemma of not being able to assess participants more than once as a result of their previous exposure. Additionally, the concentrated use of this language pair in research constraints the generalizability of findings to other foreign languages (Grimaldi et al., 2010). Finally, Nelson and Dunlosky's (1994) exclusive focus on measures of retrieval accuracy do not give a holistic indication as to the relative item difficulty of each

* Corresponding author. School of Psychology, Keynes College, University of Kent, Canterbury, CT2 7NP, UK.
E-mail addresses: wesleypyke@gmail.com (W. Pyke), a.h.javadi@gmail.com (A.-H. Javadi).

https://doi.org/10.1016/j.ssaho.2023.100398
Received 4 July 2022; Received in revised form 26 November 2022; Accepted 3 January 2023
Available online 6 January 2023

paired-associate (Grimaldi et al., 2010).

To address the criticisms of (Nelson & Dunlosky's (1994) research and plug gaps in the literature, Grimaldi et al. (2010) provided normative data for Lithuanian-English paired associates together with metacognitive assessments of item difficulty (how difficult participants believed the pairs were to learn), measurements of error frequency, and recall latency (reaction time). Lithuanian is similar to Swahili in the sense that it also employs the standard English alphabet and is not a Romance language (prior exposure to the language is unlikely), making it a logical choice to extend normative data beyond Nelson and Dunlosky's (1994) Swahili-English paired associates. Grimaldi et al.'s (2010) metacognitive assessments included: ease-of-learning judgements (EOL), which were made via the use of a scale ranging from "Very Hard" on the left side to "Very Easy" on the right side and judgements of learning (JOL), where the input consisted of a continuous percentage scale ranging from 0% "Likely To Recall" to 100% "Likely To Recall". In assessing participants' perceptions of how easy it was to learn each item (EOL) and how likely they would hypothetically be to recall any one item should they be tested one week later (JOL), Grimaldi et al. (2010) provided a subjective and therefore more thorough insight into item difficulty than Nelson and Dunlosky (1994). Furthermore, Grimaldi et al. (2010) explored the frequency of commission (false positives) and omission errors (false negatives) in addition to recall latency, which was defined as the time in seconds between the presentation of a Lithuanian word and the first keypress of the participant. In summary, in their study, Grimaldi et al. (2010) succeeded in both addressing the shortcomings of Nelson and Dunlosky's (1994) research whilst also providing normative paired-associate data for a distinct yet comparable language pair.

Nonetheless, an enduring limitation of the literature on foreign language-English paired associates is that until recently, research has been confined to the production of normative data for alphabetic writing systems such as English, Swahili, Lithuanian, and Portuguese (Grimaldi et al., 2010; Lima & Buratto, 2021; Nelson & Dunlosky, 1994), leaving languages based on logographic writing systems such as Chinese and Japanese unexplored. To address this unexplored area in the literature, Cho et al. (2020) produced normative data for 160 Chinese-English noun paired associates. Their study included measures matching those of Grimaldi et al. (2010), namely recall accuracy, recall latency (reaction time), and metacognitive EOL and JOL judgements. Importantly, the study of Chinese as a language is intriguing and useful for reasons beyond its logographic nature. Chinese has unique characteristics in the sense that approximately 90% of characters consist of sub-characters (i.e., radicals), which provide an underlying indication as to the meaning or the pronunciation of the entire character (Hoosain, 1991; Cho et al., 2020). Additionally, more visually complex Chinese characters take up an identical amount of space as less visually complex characters, providing an inherent level of experimental control and comparability when studying item difficulty (Cho et al., 2020).

Although Cho et al.'s (2020) investigation of an idiographic language system is undoubtedly valuable, it is intuitive to apply Grimaldi et al.'s (2010) earlier criticism of Nelson and Dunlosky's (1994) study of Swahili-English word pairs here: where possible, further language pairs should be investigated to increase the generalizability of findings and to overcome the issue of limited sample sizes. Furthermore, in all previous normative studies, whether they assess lexical or logographic difficulty, the response item has always been in the participant's native language. This undoubtedly draws upon previous semantic representations which will vary from person to person.

The current study provides three normative data sets for an additional language, namely Japanese, and presents a uniquely comprehensive set of results through the implementation of three distinct conditions (logographic, lexical and abstract) designed to both extend previous research and provide novel insights. In our logographic condition, we explored Japanese symbol (kanji) to English word paired associates to assess whether Cho et al.'s (2020) findings extended to the Japanese language. In our lexical condition, we performed a novel investigation of Japanese word to English word pairs. The Japanese words in this condition represented the expression of the Japanese language in Roman script (romaji). Finally, we created an abstract condition that involved Japanese symbol - Japanese word paired associates. This gave us insight into how participants would perform in terms of recall accuracy should they be unable to draw on any previous semantic representations. To elaborate, as participants have not been exposed to either Japanese symbols or words prior to measurement, we presumed that they were unable to draw on existing anchor points (i.e. attempting to visualise the English word in the symbol). The enhanced difficulty of this condition was validated in our pilot study. Indeed, participants in this abstract group exhibited significantly lower recall accuracy than that of participants in the logographic and lexical conditions where subjects could lean on the familiar English word pair components. The primary aim of this study was to provide a complete normative data set that experimenters could customize as needed. To this end, the inclusion of an abstract condition minimized the influence of several variables (such as prior knowledge) that could have a confounding effect on data produced by various interventions (e.g., pharmacological, brain stimulation etc).

The secondary aim of this study was to carry out an intergroup comparison of both logographic and lexical conditions by means of a *t*-test. This was carried out to establish variations in difficulty between logographic and lexical stimuli. Research demonstrates that pictures are remembered better than words (Grady et al., 1998; Paivio, 1971; Shepard, 1967; Standing et al., 1970). Whilst logographic symbols resemble pictures, in the sense that both are visual depictions, they cannot be directly compared as symbols carry far fewer variables than pictures (e.g., one colour, one size, similar in shape). However, based on the previous literature comparing pictures and words, we hypothesized that recall accuracy for logographic stimuli would be significantly higher than lexical.

2. Methods

2.1. Participants

Initially, 240 participants signed up for the study, of which 189 were recruited through an online survey platform (Prolific) and 51 were psychology undergraduates from the University of Kent. The data of 7 participants were excluded due to inattentiveness or exceptionally poor accuracy (remembering less than 2 pairs in the final testing session). The study information sheet and consent form stated that, in order to be eligible, participants had to confirm that they had no previous knowledge of logographic language systems such as Korean, Japanese, or Chinese. The mean age was 21.8 ($SD = 3.6$), and 71% of participants were female. Ethical approval was obtained from the University of Kent ethics board.

2.2. Materials

To obtain our stimuli, we randomly chose 10 Japanese symbols/words (Kanji/Romaji) and their English equivalents from 12 categories on the following educational website: https://www.learn-japanese-adventure.com/japanese-words.html. The only criterion was that the words had only one Kanji (symbol) and one Romaji (word) expression. This totalled 120 triads (Japanese symbol - English word, Japanese word - English word and Japanese symbol - Japanese word). 10 triads were deducted from this total as they were employed in our pilot study. This experiment included the remaining 110 triads (330 pairs). To obtain the English and Japanese word frequency values, we used the following GitHub repository https://github.com/hermitdave/FrequencyWords. The mean stroke count of Japanese symbols and mean length of Japanese words and English words can be seen in Table 1.

Table 1
Lexical characteristics of English words, Japanese words, and Japanese symbols (Mean).

	Stroke Count	Length
Japanese symbols	12 (SD = 4.78)	–
Japanese words	–	4 (SD = 1.34)
English words	–	5 (SD = 1.58)

Notes: Length: number of characters.

2.3. Procedure

Participants were randomly assigned to one of three conditions, each of which presented one combination of paired associates: Japanese symbols - English words (Logographic), Japanese words - English words (Lexical), or Japanese symbols - Japanese words (Abstract). Both the Logographic and Lexical conditions each contained 25 paired associates, the Abstract contained only 15, as our pilot study indicated that participants performed most poorly in the Abstract condition, and we sought to prevent a floor effect. Participants were told that they would study either Japanese/English or Japanese/Japanese pairs and that their knowledge would subsequently be tested. Additionally, participants completed ease-of-learning (EOL) and judgements-of-learning (JOL) measures, which were preceded by program instructions. Upon initiation of the experiment, participants entered a practice phase, upon which they studied 10 paired associates, the nature of which depended on the condition. A short test was administered after this practice round in which participants had to fill out the correct paired associate. Each paired associate was presented for 9.5 s before automatic progression to the next pair.

After completing the practice phase, an on-screen indication was given that the actual experiment would begin. Depending on the condition, either 25 (lexical and logographic) or 15 (abstract) paired associates were randomly selected from the pool of 110 stimuli. Each pair was presented to the participant in the study phase for a duration of 9.5 s, after which the program automatically presented the next pair. The pairs were presented in a randomized order and shown as black text/symbol on a white background, with one half of the pair to the left of the screen and the other on the right (see Fig. 1). After the study phase, the participant entered a test phase in which they were prompted to enter the missing paired-associate (all studied pairs were presented) and press the enter key to proceed. The participant had unlimited time to enter the answer, after which the next question was presented automatically. Each participant then repeated this process twice more for a total of three study-test cycles.

Following completion of the final cycle, participants were instructed on-screen to complete the metacognitive judgements for each paired associate they had learnt. The Ease of Learning measure presented a paired associate and asked participants the question: "How difficult was it to remember the following translation?". On the same screen, participants were instructed to indicate their judgement on a continuous scale ranging from 0, which was labelled "Very Easy", to 100, which was denoted as "Very Hard". The Judgement of Learning measure was similar in format, with the question: "How likely is it that you will remember the following translation in one week's time?" and the 0 to 100 scale presenting the terms "Very Unlikely" on the left and "Very Likely" to the right. Both scales presented anchor points at 0, 25, 50, 75 and 100 to encourage full utilisation of the scale's range. No time limit was imposed for either metacognitive judgement task. The total duration of the study was 30 min.

2.4. Marking criteria

Certain criteria were considered when marking the answers of participants as correct or incorrect. Minor spelling errors, such as missing letters (e.g., writing '*fores*' in place of '*forest*'), abbreviations (e.g., writing '*grandma*' instead of '*grandmother*'), and typos (e.g., '*shik*' instead of '*ship*') were scored as correct. Typos were only marked as incorrect if they produced another word (e.g., '*mountain*' instead of the correct answer of '*fountain*'). The marker, who was the same throughout, was more lenient when marking answers in the abstract (Japanese symbol - Japanese word) condition, to consider the pronounced difficulty of this condition.

2.5. Statistical analysis

Our statistical analysis of the data involved several tests. We performed a Mixed ANOVA to assess whether the differences in terms of mean recall accuracy across trials in each condition were statistically significant and calculated partial eta squared (η_p^2) to indicate effect sizes. Pearson's r correlations were calculated to give insight into the relative item difficulty across trials in each condition. The simultaneous multiple regressions allowed us to assess whether - and the extent to which - stroke count and other lexical characteristics relevant to different conditions were predictive of recall accuracy. Mean ratings on the metacognitive EOL and JOL judgement scales gave a general indication as to the perceived difficulty of the pairs in each condition, whilst Pearson's r correlations showed the degree and direction to which these ratings were associated with participants' actual recall accuracy. Importantly, the EOL scale was reversed compared to the JOL rating scale, which should be considered when analysing results. A post-hoc independent samples t-test gave insight into whether there were any significant differences in mean recall accuracy between conditions. Cohen's (1988) benchmarks for effect sizes were utilised throughout this study.

3. Results

A mixed ANOVA with Trial (Accuracy 1, 2 and 3) as the within subjects factor and condition (Logographic, Lexical and Abstract) as the between group factor was run. There was a significant main effect of

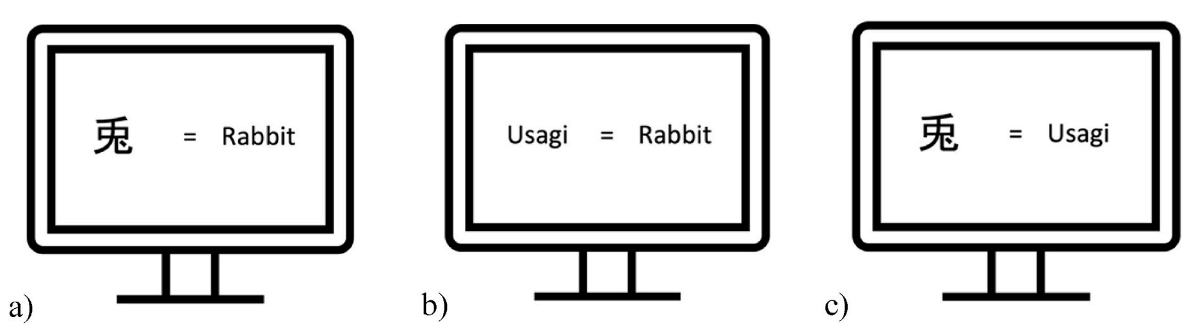

Fig. 1. An example of how each condition was displayed to participants: a) Japanese Symbol – English Word; b) Japanese Word – English Word; c) Japanese Symbol – Japanese Word. Each participant was allocated to one condition only with 25 (for logographic and lexical) and 15 (abstract) pairs randomly chosen from a stimuli bank of 110 triads.

Trial, $F(1.799, 588.327) = 1002.539$, $p < .001$, $\eta_p^2 = 0.754$. There was also a significant main effect of Condition, $F(2, 327) = 65.442$, $p < .001$, $\eta_p^2 = 0.286$. The interaction effect (Trial*Condition) was small, although significant $F(3.598, 588.327) = 10.382$, $p < .001$, $\eta_p^2 = 0.060$.

Due to the significant main effect of Condition, a post hoc comparison was run.

3.1. Post-hoc independent samples T-test

Three independent groups *t*-test were performed to compare mean recall accuracy between the three conditions (Logographic, Lexical and Abstract). The results of the t-tests are compiled in Table 2. Contrary to our hypothesis that symbols would be remembered better than words, mean recall accuracy for the lexical condition was significantly higher.

3.2. Japanese symbol - English word (logographic)

Recall accuracy increased across trials in the Japanese Symbol - English Word condition: trial 1 = 32.3 (*SD* = 19.2), trial 2 = 61.0 (*SD* = 17.1), trial 3 = 74.6 (*SD* = 15.4). Relative difficulty of the items was similar across trials as shown by the equally strong positive correlations between trial 1 and trial 2, $r(108) = 0.647$, $p < .001$, and trial 2 and trial 3, $r(108) = 0.720$, $p < .001$. The relative influence of lexical and logographic characteristics on the difficulty of pairs was assessed by means of a simultaneous multiple regression for each trial in each condition using the relevant predictor variables. The predictor variables for the Japanese symbol - English word condition were: English word frequency, English word length, and Japanese symbol stroke count. Together, these factors accounted for 18.8%, 14.3%, and 4.5% of the variance in trials 1–3, respectively. However, only the Japanese symbol stroke count proved to be a reliable and significant predictor, which can be seen in Table 3.

The mean EOL rating for this condition was 49.56 (*SD* = 10.89), indicating that participants perceived the pairs to be moderately difficult to learn. A moderate and significant negative correlation was found between EOL and recall accuracy on the third trial (see Table 4). This correlation meant that those pairs rated as more difficult to remember (higher EOL score) were associated with lower recall accuracy. Put differently, pairs rated as easier to remember (lower EOL score) were associated with higher recall accuracy. The mean JOL value for this condition was 41.10 (*SD* = 10.72). This indicates that participants were moderately confident that they could recall the pairs a week later. A significant, moderate positive correlation was identified between JOL and recall accuracy on the third trial (see Table 4), such that pairs that participants felt they could recall more easily in a week's time (higher JOL score) were associated with higher recall accuracy. Vice-versa, pairs that participants felt they would be less able to recall in a week's time (lower JOL score) were linked to lower recall accuracy. For all three conditions, the correlations between metacognitive judgements and accuracy on the third trial are exhibited in Table 4.

Table 2
Mean recall accuracy across conditions and Post hoc independent t-tests comparing the mean recall accuracy between the three conditions.

Condition	Mean	SD	t	df	p	Cohen's D
Logographic vs Lexical	56.0 64.4	14.8 14.7	−4.242	218	< **.001***	0.6
Logographic vs Abstract	56.0 40.3	14.8 17.3	7.078	218	< **.001***	1.0
Lexical vs Abstract	64.4 40.3	14.7 17.8	10.941	218	< **.001***	1.5

Note. *$p < .005$.

3.3. Japanese word - English word (lexical)

Recall accuracy again increased across test trials in the Japanese word - English word condition: trial 1 = 45.3 (*SD* = 18.5), trial 2 = 67.9 (*SD* = 18.5), trial 3 = 80.1 (*SD* = 14.3). As in the Japanese symbol - English word condition, the relative difficulty of the items remained similar across trials in the Japanese word - English word condition, as shown by strong, significant, and positive correlations between trials 1 and 2, $r(108) = 0.647$, $p < .001$ and trials 2 and 3, $r(218) = 0.720$, $p < .001$. Regarding the multiple regression analysis, the predictor variables for this condition included: frequency of English words, English word number of letters, and Japanese word number of letters. These factors accounted for 0.7%, 4.6%, and 3.2% of the variance in trials 1–3, respectively. However, these predictors were generally non-significant and therefore unreliable, as shown in Table 3.

The mean EOL rating for this condition was 50.72 (*SD* = 13.34), suggesting that participants perceived the pairs to be moderately difficult to learn. A strong, significant negative correlation was found between EOL ratings and recall accuracy on the third trial (see Table 4), indicating that lower recall accuracy scores were associated with those pairs rated as being more difficult to remember (higher EOL score). Vice-versa, higher recall accuracy scores were linked to pairs rated as easier to remember (lower EOL score). The mean JOL rating for this condition was 44.38 (*SD* = 11.45), indicating that participants were moderately confident that they could recall any pair a week later. A moderate, significant positive correlation was identified between JOL scores and recall accuracy on the third trial (see Table 4).

3.4. Japanese symbol - Japanese word (abstract)

Finally, accuracy again increased across trials in the Japanese symbol - Japanese word condition: trial 1 = 23.9 (*SD* = 19.1), trial 2 = 41.2 (*SD* = 19.4), trial 3 = 55.8 (*SD* = 19.9). As in the other conditions, relative difficulty of the items remained similar across trials in the Japanese word - Japanese symbol condition, as is evidenced by the strong positive correlations in terms of recall accuracy between trial 1 and trial 2, $r(108) = 0.742$, $p < .001$ and trial 2 and Trial 3, $r(108) = 0.817$, $p < .001$. In terms of the multiple regression analysis, the predictor variables for this condition were: Japanese word length and Japanese symbol stroke, which accounted for 29.5%, 19.2%, and 19.4% of the variance in trials 1–3, respectively. Both Japanese symbol stroke count and Japanese word length proved to be highly reliable and significant predictors of recall accuracy. This can be seen in Table 3.

The mean EOL value for this condition was 58.68 (*SD* = 13.69), indicating that participants perceived these pairs to be more difficult to learn than those in the Japanese symbol - English word (*M* = 49.56, *SD* = 10.89) and Japanese word - English word (*M* = 50.72, *SD* = 13.34) conditions. A strong, significant negative correlation was found between EOL and recall accuracy in the third trial (see Table 4). Those pairs rated as more difficult to remember (higher EOL score) were associated with lower recall accuracy whereas pairs rated as easier to remember (lower EOL score) were correlated with higher recall accuracy. The mean JOL value for this condition was 35.42 (*SD* = 12.16). This score was lower than in the Japanese symbol - English word (*M* = 41.10, *SD* = 10.72) and Japanese word - English word (*M* = 44.38, *SD* = 11.45) conditions, indicating that participants in this group were the least confident that they could recall the pairs a week later. A significant, moderate positive correlation was identified between JOL and recall accuracy on the third trial (see Table 4), suggesting that pairs that participants felt they could recall more easily in a week's time (higher JOL score) were associated with higher recall accuracy. Vice-versa, pairs that participants felt they would be less able to recall in a week's time (lower JOL score) were linked to lower recall accuracy.

Table 3
Lexical characteristics relevant to each condition predicting accuracy across trials.

Condition	Predictors	Test Trial								
		1			2			3		
		Std. β	*t*	*p*	Std. β	*t*	*p*	Std. β	*t*	*p*
Logographic	Eng. Word Freq.	0.231	2.561	**.012***	0.113	1.220	.225	0.051	0.525	.601
	Eng. Word Len.	0.120	1.366	.175	0.019	0.206	.837	−0.055	−0.578	.564
	Jap. Sym. Stroke Count.	−0.309	−3.421	< **.001***	−0.336	−3.622	< **.001***	−0.183	−1.869	.064
Lexical	Eng. Word Freq.	−.050	−0.515	.608	0.022	0.226	.822	0.129	1.339	.183
	Eng. Word Len.	−0.045	−0.459	.647	−0.202	−2.103	.038	−0.119	−1.237	.219
	Jap. Word Len.	0.053	0.543	.588	−0.045	−0.470	.639	0.065	0.670	.504
Abstract	Jap. Sym. Stroke Count.	−0.314	−3.822	< **.001***	−0.359	−4.082	< **.001***	−0.321	−3.653	< **.001***
	Jap. Word Len.	−0.398	−4.847	< **.001***	−0.201	−2.287	.024	−0.257	−2.922	**.004***

Note. Std. β: Standardized betas employ standard deviations as units (Z-scores), which allows for a fair and valid comparison of predictor effects. Eng.: English, Jap.: Japanese, Freq.: Frequency, Len.: Length. Logographic: Japanese symbol - English word paired associates, Lexical: Japanese word - English word paired associates, Abstract: Japanese symbol - Japanese word paired associates. A False Discovery Rate threshold of 0.017 was used for the purpose of correction for multiple comparisons (Benjamini & Hochberg, 1995).

Table 4
Correlations between metacognitive judgements and third trial accuracy by condition.

Condition	Judgements	Pearson's r	p
Logographic	EOL	−0.484	< **.001***
	JOL	0.440	< **.001***
Lexical	EOL	−0.614	< **.001***
	JOL	0.347	< **.001***
Abstract	EOL	−0.787	< **.001***
	JOL	0.694	< **.001***

Note. Degrees of freedom for all reported correlations equals 108. EOL: ease-of-learning judgements, JOL: judgements of learning, Logographic: Japanese symbol - English word paired associates, Lexical: Japanese word - English word paired associates, Abstract: Japanese symbol - Japanese word paired associates. *$p < .005$.

4. Discussion

The main purpose of this study was to provide normative paired-associate data for both logographic (Japanese symbol - English word), lexical (Japanese word - English word), and abstract pairs (Japanese symbol - Japanese word) derived from a previously unexplored logographic language system (Japanese). The primary measure in this experiment was recall accuracy, indicating item difficulty, and was found to increase across trials in all conditions. Relative item difficulty, which was assessed through correlating recall accuracy in trials of the same condition, remained similar. A simultaneous multiple regression highlighted that symbol stroke count was the only significant and reliable predictor of recall accuracy. We correlated participants' metacognitive judgements (perceptions) of pair difficulty with their recall accuracy and discovered that subjects were aware of how well they had learned the pairs. Finally, an intergroup comparison of mean recall accuracy across the three conditions by means of a *t*-test revealed significant differences. Specifically, recall accuracy was higher in the lexical condition than in the logographic group.

Two common trends were revealed across conditions. Indeed, we found that recall accuracy increased with repeated stimuli exposure (trials), as was the case in previous studies seeking to provide normative data for foreign language (either words or symbols) and English word paired associates (Nelson & Dunlosky, 1994; Grimaldi et al., 2010; Cho. et al., 2020; Lima & Buratto, 2021). This pattern was to be expected as participants' repeated exposure to the paired associates produced learning and enhanced memorisation. As in previous studies (Nelson & Dunlosky, 1994; Grimaldi et al., 2010; Cho. et al., 2020), the relative difficulty of items across trials remained similar. This was important to measure as pairs were randomised between trials, such that certain combinations - which may have been easier or more difficult - could have confounded recall accuracy. The strong correlations that we identified between recall accuracy in all three trials of all conditions alleviated these concerns.

In terms of the simultaneous multiple regression, Japanese symbol stroke count was found to be a significant predictor of recall accuracy in five of six trials in the logographic and abstract conditions, such that more complex symbols (higher stroke count) were associated with lower recall accuracy, and vice-versa. None of the other predictors were found to be consistently significant. In particular, English word frequency was found to not be a significant predictor of recall accuracy, which is surprising if we consider that earlier studies into the norms of paired associates derived from alphabetic language systems revealed significant positive albeit weak correlations between this variable and recall accuracy (Grimaldi et al., 2010; Nelson & Dunlosky, 1994). However, in contrast, Cho et al.'s (2020) investigation into the paired-associate norms of a logographic language system (Chinese) actually mirrored our results, such that stroke count was the only significant predictor of recall accuracy whereas English word frequency was not significant throughout. Arguably, for paired associates derived from logographic language systems, stimuli complexity (stroke count) overwhelms the influence of any other predictors, and it may be the only variable that contributes significantly to the latent variable of item difficulty (which recall accuracy indicated). We also posit that, as Nelson and Dunlosky (1994) and Grimaldi et al. (2010) did not employ a regression analysis, cause and effect cannot be assumed and the correlations these researchers identified may be spurious in nature. Finally, we found that the proportion of variance accounted for by the predictors in the Japanese symbol and English word condition dropped off significantly across trials whilst remaining steady in Cho et al.'s (2020) research. This reduced contribution of the predictors to the model in this condition could signal that, in our study, participants' individualised learning techniques began to account for a progressively larger proportion of the variance in recall accuracy as trials continued.

Across conditions, positive, significant, and moderate to strong correlations were identified between EOL judgements and recall accuracy. This relationship was strongest in the abstract condition and weakest in the logographic condition. We could speculate that participants' familiarity with the English word components in the latter condition may have introduced a certain bias that led subjects to underestimate the difficulty of recalling pairs whereas the complete novelty of stimuli in the former condition did not. A similar pattern was observed in regard to the relationship between JOL ratings and recall accuracy, which further reinforces the existence of such a perceptive bias. Indeed, significant, strong and moderate positive correlations were revealed between JOL ratings and recall accuracy in the abstract and logographic conditions, respectively. This correlation was not significant in the lexical condition. Overall, it seems that participants seemed to be aware of how successfully they had learned the material which was demonstrated by the close alignment between the perceived difficulty of the pairs in hindsight and actual recall. In addition to these measures, it would have also been useful to test recall accuracy one week later in order to establish whether the participants perceived JOL aligned with actual recall accuracy. This would help not only further inform metacognitive accuracy, but also provide a greater understanding of the normative dataset.

Following a significant between subjects main effect of condition, a post hoc independent samples *t*-test revealed significant differences between all three groups, with lexical stimuli more easily remembered than logographic. These results are unexpected if we consider that humans typically remember pictures far better than words (Shepard, 1967; Standing et al., 1970; Paivio, 1971; Grady et al., 1998). However, there are differences between pictures and the logographic symbols employed in our study, which may have influenced our findings. Specifically, our symbols were somewhat standardized in the sense that they were coloured black and white and occupied the exact same amount of space. As these attributes are not characteristic of typical pictures which are diverse and unstructured, we could infer that, in our study, participants could not draw on existing semantic representations to aid memorisation - the abstract nature of the stimuli also controlled for the confounding influence of prior exposure. Accordingly, we could speculate that as our participants all had no prior knowledge of logographic writing systems, they were more familiar with the Romanised equivalent (Romaji), allowing them to use similar encoding techniques to their native languages.

5. Conclusion

Our provision of novel and diverse normative data, which detailed the specific characteristics of the stimuli set (stroke count, word length, etc.), allows researchers to customize stimuli composition, manipulate it as they see fit, and counterbalance it appropriately in future studies. Besides this potential for widespread application, we demonstrated that stimuli complexity (stroke count) was the only significant predictor of recall accuracy, which highlights the importance of this variable in memory research. In addition, we found that lexical paired associates were remembered better than logographic equivalents, which contradicts previous findings regarding picture-word memory. The normative data that we provided can be used to explore how to optimize language learning or refine existing models of memory.

Ethics approval

Ethical approval was obtained from the University of Kent ethics board.

Funding

No funding was provided for this research.

Consent to participate

Informed consent was obtained from all individual participants included in the study.

Consent for publication

The authors give consent for the journal to publish the current manuscript.

Code availability

N/A.

Availability of data and materials

Stimuli and data for this study can be downloaded from https://doi.org/10.17605/OSF.IO/AHXQ4.

CRediT authorship contribution statement

Wesley Pyke: Conceptualization, Resources, Investigation, Formal analysis, Writing – original draft, Writing – review & editing. **Johan Lunau:** Investigation, Writing – original draft. **Amir-Homayoun Javadi:** Conceptualization, Methodology, Visualization, Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ssaho.2023.100398.

References

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B, 57*(1), 289–300.

Cho, K. W., Tse, C. S., & Chan, Y. L. (2020). Normative data for Chinese-English paired associates. *Behaviour Research Methods, 52*(1), 440–445. https://doi.org/10.3758/s13428-019-01240-2

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale: NJ Lawrence Erlbaum Associates, Publishers. - References - Scientific Research Publishing. Cohen, J. (1988), (New York: Routledge Academic).

Grady, C. L., McIntosh, A. R., Rajah, M. N., & Craik, F. I. M. (1998). Neural correlates of the episodic encoding of pictures and words. *Proceedings of the National Academy of Sciences, 95*(5), 2703–2708. https://doi.org/10.1073/PNAS.95.5.2703

Grimaldi, P. J., Pyc, M. A., & Rawson, K. A. (2010). Normative multitrial recall performance, metacognitive judgments, and retrieval latencies for Lithuanian-English paired associates. *Behavior Research Methods, 42*(3), 634–642. https://doi.org/10.3758/BRM.42.3.634

Hoosain, R. (1991). *Psycholinguistic implications for linguistic relativity: A case study of Chinese*. PsycNET.

Jang, Y., & Nelson, T. O. (2005). How many dimensions underlie judgments of learning and recall? Evidence from state-trace methodology. *Journal of Experimental Psychology: General, 134*(3), 308–326. https://doi.org/10.1037/0096-3445.134.3.308

Kelemen, W. L., Winningham, R. G., & Weaver, C. A. (2007). Repeated testing sessions and scholastic aptitude in college students' metacognitive accuracy. *European Journal of Cognitive Psychology, 19*(4–5), 689–717. https://doi.org/10.1080/09541440701326170

Lima, M. F. R. De, & Buratto, L. G. (2021). Norms for familiarity, concreteness, valence, arousal, wordlikeness, and recall accuracy for Swahili–Portuguese word pairs. *Sage Open, 11*(1), Article 215824402098852. https://doi.org/10.1177/2158244020988524

Nelson, T. O., & Dunlosky, J. (1994). Norms of paired-associate recall during multitrial learning of Swahili-English translation equivalents. *Memory, 2*(3), 325–335. https://doi.org/10.1080/09658219408258951

Paivio, A. (1971). *Imagery and verbal processes*. https://doi.org/10.4324/9781315798868. Imagery and Verbal Processes.

Richards, R. M., & Nelson, T. O. (2004). Effect of the difficulty of prior items on the magnitude of judgments of learning for subsequent items. *American Journal of Psychology, 117*(1), 81–91. https://doi.org/10.2307/1423597

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior, 6*(1), 156–163. https://doi.org/10.1016/S0022-5371(67)80067-7

Standing, L., Conezio, J., & Haber, R. N. (1970). Perception and memory for pictures: Single-trial learning of 2500 visual stimuli. *Psychonomic Science, 19*(2), 73–74. https://doi.org/10.3758/BF03337426

Toppino, T. C., & Cohen, M. S. (2009). The testing effect and the retention interval questions and answers. *Experimental Psychology, 56*(4), 252–257. https://doi.org/10.1027/1618-3169.56.4.252

Van Overschelde, J. P., & Nelson, T. O. (2006). Delayed judgments of learning cause both a decrease in absolute accuracy (calibration) and an increase in relative accuracy (resolution). *Memory & Cognition, 34*(7), 1527–1538. https://doi.org/10.3758/BF03195916