Title

Standardizing norms for 1286 colored pictures in Cantonese

Authors

Jing Zhong[1,2+], Weike Huang[1,2+], Keyi Kang[1,2], Jon Andoni Duñabeitia[3,4], Christos Pliatsikas[5], Haoyun Zhang[1,2*]

[1]Centre for Cognitive and Brain Sciences and [2]Department of Psychology, University of Macau, Taipa, Macau SAR, China.

[3]Centro de Investigación Nebrija en Cognición (CINC), Universidad Nebrija, Madrid, Spain.

[4]UiT The Arctic University of Norway, Tromsø, Norway

[5]School of Psychology and Clinical Language Sciences, University of Reading, Reading, UK

[+]The two authors contribute equally to this project.

Address for correspondence

Haoyun Zhang, Ph.D.

Centre for Cognitive and Brain Sciences, ICI

University of Macau, N21-1015f, Taipa, Macau SAR, China.

Telephone number: (853) 88229212

Email: haoyunzhang@um.edu.mo

Abstract

This study established psycholinguistic norms in Cantonese for a set of 1286 colored pictures sourced from several picture databases, including 750 colored line drawings from MultiPic (Duñabeitia et al., 2018) and 536 photographs selected for McRae et al. (2005) concepts. The pictures have undergone rigorous normalization processes. We provided picture characteristics such as name and concept agreement, familiarity, visual complexity, as well as the frequency of modal responses. Through correlational analyses, we have observed strong interrelationships among these variables. We also compared the current Cantonese norming to other languages and demonstrated similarity and variations among different languages. Additionally, we embraced the multilingual diversity within the current sample, and found that higher Cantonese proficiency but lower non-native language proficiency was associated with better spoken picture naming. Last but not least, we validated the predictivity power of normed variables calculated from typing responses to spoken picture naming, and the consistency between typing and spoken responses. The present norming provides a timely and valuable alternative for researchers in the field of psycholinguistics, especially those studying Cantonese production and lexical retrieval. All raw data, analysis scripts, and final norming results are available online as psycholinguistic norms for Cantonese in the following link at Open Science Framework, https://osf.io/dz9j6/?view_only=a452d8a56c92430b9dedf21ac26b1bc1.

Keywords Cantonese, picture norm, cross-language, name agreement, reaction time

1. Introduction

Lexical retrieval plays a pivotal role in human communication and has long been a focal point of psycholinguistic research. Within the vast body of studies exploring this topic, picture naming has been one of the most extensively used paradigms. During picture naming, participants are typically presented with a picture and asked to provide one precise response to describe the name of the picture. By examining the time it takes individuals to generate the response, and the accuracy of the response, researchers can gain insights into the cognitive processes underlying language production.

Many picture characteristics and psycholinguistic variables have been demonstrated to affect the naming process. For instance, one picture could have one or many names, and the concept of name agreement describes to what extent people agree on the name of the picture. Many studies have shown that lower name agreement was associated with slower response time with lower accuracy (Alario et al., 2004; Dell'acqua et al., 2000; Liu et al., 2011; Perret & Bonin, 2019; Pind & Tryggvadóttir, 2002; Rodríguez-Ferreiro et al., 2009; Valente et al., 2014), because of the higher selection demand for pictures with more labels during name retrieval. In addition, image familiarity is generally assumed to affect the ease of semantic representation with higher familiarity being associated with shorter naming latency (Brysbaert et al., 2014; Kremin et al., 2001; Rodríguez-Ferreiro et al., 2009; Zannino et al., 2010). Furthermore, visual complexity, which represents the intricacy of depicted objects, has also been shown to affect picture naming performance such that a more complex picture is more difficult to comprehend, leading to longer name latencies (Alario et al., 2004), although it is not entirely consistent (Ellis & Morrison, 1998). In addition to picture properties, psycholinguistic variables for the modal names of pictures would also affect picture naming performance. For example, higher word frequency and shorter word length have been shown to be related to faster response time (Brysbaert et al., 2014; Gollan et al., 2008; LaGrone & Spieler, 2006; Newman & German, 2005). In addition, the age of acquisition (AoA) has also been shown to affect picture naming though the effect is not always consistent, because it confounds largely with word frequency (Brysbaert et al., 2014). Given the potential effects of these pictures and psycholinguistic characteristics on picture naming, it is crucial for researchers to be aware of these variables when selecting pictures for their studies. Consequently, the images used in picture naming research must undergo a process of normalization.

Normalization entails having a representative group of individuals assess the images based on various criteria that describe their fundamental features. This systematic approach ensures that the selected images are standardized and meet the necessary requirements, thereby guaranteeing precise and reliable research outcomes. As early as 1980, a set of 260 black-and-white line drawings was standardized in English (Snodgrass & Vanderwart, 1980), cited thousands of times, and extended to multiple languages (Bangalore et al., 2022; Cuetos et al., 1999; Pind et al., 2000; Tsaparina et al., 2011; Weekes et al., 2007). In this database, factors such as name agreement, familiarity, visual complexity and image agreement were reported for each picture. Since then, numerous norms have been developed, each offering a distinct collection of pictures along with their normative properties (Adlington et al., 2009; Barbarotto et al., 2002; Bonin et al., 2003; Brodeur et al., 2010; Cuetos & Alija, 2003; Dan-Glauser & Scherer, 2011; De Winter & Wagemans, 2004; Dell'acqua et al., 2000; Francisco J. Moreno-Martínez & Montoro, 2012; Janssen et al., 2011; Kremin et al., 2003; Magnié et al., 2003; Migo et al., 2013; Nishimoto et al., 2005; Nishimoto et al., 2010; Rossion & Pourtois, 2004; Salmon et al., 2010; Viggiano et al., 2004). Overall, the availability of these diverse norms has significantly enhanced the research landscape, enabling researchers to access standardized and extensively evaluate stimuli across various languages, thereby facilitating rigorous cross-language investigations.

Across languages, the properties of pictures and psycholinguistic variables can vary. For example, the discrepancy in name agreement is evident when comparing the same 500 pictures in the MultiPic dataset (Duñabeitia et al., 2022), where the mean H statistics (an index for name agreement) was 0.30 in Spanish and 1.07 in Chinese. Such differences are to be expected, as each language possesses its unique lexicon and linguistic conventions. Moreover, the impact of the above-mentioned factors on picture naming performance may not always be consistent across languages. For example, studies have demonstrated that image familiarity exerts a more pronounced influence on picture naming in Mandarin Chinese (Chen & Zhu, 2015; Zhou & Chen, 2017) compared to other languages such as Japanese (Nishimoto et al., 2005), Italian (Dell'acqua et al., 2000) or French (Bonin et al., 2003). Because of these cross-language variances, it is imperative to have available norms specifically tailored to the target language of interest to researchers, which ensures that the chosen stimuli accurately reflect the linguistic and cultural characteristics of the particular language, thus enabling more precise and meaningful investigations.

Nevertheless, the majority of the norms are only available in English, and some have been extended to European languages, but only a few are available in Asian languages such as a Japanese normative set containing 359 pictures (Nishimoto et al., 2005), and a Korean version of the Boston naming test brought out by Kim and Na (1999). As a large-capacity language, Chinese embraces multiple branching languages like Mandarin and Cantonese, which do not yet have much norming available. To the best of our knowledge, there are only a few published norms available in Mandarin Chinese (Chen & Zhu, 2015; Liu et al., 2011; Momenian et al., 2021; Ni et al., 2019; Weekes et al., 2007; Zhou & Chen, 2017), along with norms in other formats such as typing (Wang et al., 2020). Yet, there are only two norms in Cantonese with limited number of pictures (Momenian et al., 2021; Momenian et al., 2022).

Cantonese has been applied as a lingua franca among residents in Southern China including Hong Kong and Macao SAR, which has developed its own written forms for massive public communication (Li, 2006). According to the 2018 Thematic Household Survey (THS) on the use of language in Hong Kong, a remarkable 87.7% of residents aged 6 to 65, with normal hearing and speech abilities, rated their language competence in Cantonese as very good or good (Census and Statistics Department of Hong Kong, 2020). This finding highlights the widespread application and acceptance of Cantonese within the local community. Furthermore, the influence of Cantonese reaches far and wide due to immigration from coastal China spanning several centuries. In Malaysia, the waves of immigration from southern China contributed to the prevalence of Cantonese in local communities. Similarly, in Singapore, the use of Cantonese remains significant, with 15.4% of the Chinese resident population, primarily those aged 65 and above, speaking Cantonese as their dialect (Fah, 2017; Vollmann & Wooi, 2018; Yu, 2022). Notably, even in Canada, the influx of Cantonese-speaking Chinese immigrants surpassed 100,000 by 1983. These examples demonstrate the enduring influence and adoption of Cantonese in various foreign communities, reinforcing its global significance. Besides, Cantonese and Mandarin are mutually unintelligible, meaning that speakers of one variety typically cannot understand the other without prior exposure or learning, because Cantonese has its own set of grammar, vocabulary, and pronunciation rules that distinguish it from Mandarin (Snow, 2008). Therefore, the norms in Mandarin Chinese cannot be directly applied to studies that target at Cantonese. Noticeably, the current availability of accessible picture databases in Cantonese falls considerably short of researchers' expectations and practical demands, as shown by the limited number of existing studies and documents. With all these reasons combined, the current study is presented, contributing to Cantonese norming by providing norms based on several established picture databases (framed as two concept sets).

The first picture norm we selected is the MultiPic norm (Duñabeitia et al., 2018). It comprises of 750 colored line drawings representing concrete concepts, which have been normed across six different European languages including British English, Spanish, French, Dutch, Italian, and German. Furthermore, a subset of 500 pictures from the MultiPic database was extended to 32 languages or language varieties (Duñabeitia et al., 2022). These pictures were commissioned by the authors and drawn by a local artist with a freehand computer application using a digital tablet and pen set. We further extended the whole MultiPic picture dataset to Cantonese norming, including the 500 pictures that have more language norms available (hereafter MultiPic500), as well as the rest of 250 pictures that only have six languages available (hereafter MultiPic250). This comprehensive resource is valuable for researchers conducting cross-language studies, providing unprecedented opportunities for comparative analysis.

Besides, we decided to expand on the McRae's semantic feature production set. This concept set, normed in English, offers a remarkable compilation of semantic features for both living and non-living concepts (McRae et al., 2005). With over 500 concepts, it stands as the largest collection of norms developed with a focus on taxonomy and semantic production. For example, for the concept *Moose*, normed features and their frequency in McRae are *is large (27), has antlers (23), has legs (14)*, and 12 other features. The McRae database also uniquely provided numerous statistics including number of features of different types for each concept, inter-correlational feature density (i.e., the relationships among feature pairs), concept similarities, along with others linguistic variables. This extensive set of norms has greatly contributed to the understanding of conceptual relationships and semantic organization. While the McRae database itself does not contain any images, its concept and feature structure provides valuable semantically related information for research on picture naming. Therefore, we sourced image materials corresponding to McRae concepts from existing databases (Brodeur et al., 2014; Moreno-Martínez & Montoro, 2012) and online resources. These photographs in the databases were directly taken by database authors and backgrounds were turned to white. All selected pictures were photographs that authentically reflect concepts provided in the McRae database. The Cantonese norming on these pictures, with a combination of the unique feature information in the McRae database would be particularly useful for psycholinguistic researchers who are interested in Cantonese-related language production and the effect of semantic structure on word retrieval.

In the current study, we first normed the total of 1286 pictures from MultiPic and McRae via typing responses and calculated numerous statistics associated with them (Experiment 1), then with a different set of participants, we demonstrated how these typing norm variables could predict the reaction time in a spoken picture naming task, which has been frequently used to measure language production (Experiment 2). Compared with previous norming, the current study has a few highlights. First, we provided norming of over one thousand pictures on Cantonese that is widely spoken but has few norming available. Second, most previous picture norming only relies on typing responses, while the current study offers not only norming based on typing, but also validates the norming with spoken picture naming reaction times.

2. Experiment 1 Picture Norming with Typing Responses

2.1 Methods

2.1.1 Participants

Eighty-one college students were first recruited to participate to norm the MultiPic picture set (Mean age = 20.7 years, SD = 2.1 years, Range =18-28 years, 43 female). Subsequently, a subset of 48 participants participated in the McRae norming portion (Mean age = 20.9 years, SD = 2.0 years, Range = 18-28 years, 24 female). All participants were recruited through advertisements posted across the university's social media. The participants all spoke Cantonese as their native and dominant language, and they were tested in Macau where Cantonese is the predominant language. Yet, all these participants were also able to speak one or more other languages. Among these participants, 83%

participants reported their second language to be Mandarin Chinese, while 73% participants reported their third language to be English. All participants rated their proficiency in each language on a 1-7 scale, which was further converted on a 0-1 scale. The rating for Cantonese (L1, Mean = 0.88, SD = 0.14) was comparable with the second language (L2, Mean = 0.82, SD = 0.14; $p > .1$), which was significantly higher than the third language (L3, Mean = 0.60, SD = 0.16; $p < .001$). Informed consent was obtained from each participant prior to the beginning of the study. The study protocols, procedures, and consent forms were approved by the Research Ethics Committee of University of Macau.

2.1.2 Materials

Two databases were normed in the current study. First, we chose the whole set of 750 pictures created by Duñabeitia et al. (2018) which consists of colored line drawings of 750 objects (MultiPic500 available in 32 languages and MultiPic250 available in 6 languages). In the current study, we contribute to the database by extending the norming of the whole set to Cantonese. Additionally, since the McRae database only contains concepts with no pictures (McRae et al., 2005), we sourced corresponding pictures from several normed picture databases (Brodeur et al., 2014; Moreno-Martínez & Montoro, 2012), as well as publicly available images from the internet (536 pictures in total). In the context of the current study, all of these pictures were grouped together with the McRae concept set for reference and analysis (i.e., these pictures from different resources are referred as McRae set in the current study).

2.1.3 Procedure

Participants were tested individually via questionnaire links hosted on Qualtrics and Gorilla for norming the picture names, and corresponding familiarities. First, a Qualtrics link containing MultiPic250 was distributed to participants. At the beginning of this portion, participants were asked to confirm again that Cantonese was their native and dominant language. At the end of the first portion, they were re-directed to the second part hosted on Gorilla which contained MultiPic500. The Gorilla portion of the 500 pictures has been translated and normed in many languages (Duñabeitia et al., 2018), and has been directly translated into Cantonese for the current study. This way, the norming for this 500-picture set was kept consistent with the norms in other languages. After participants finished the MultiPic dataset norming, they were invited to participate in the McRae dataset norming on a different day. The McRae picture norming was hosted on Qualtrics, with the same procedure as the MultiPic dataset. The entire experiment was self-paced. The MultiPic norming lasted about 90-120 minutes, and the McRae norming lasted about 30-60 minutes.

For each picture, participants were instructed to type the name as well as rating the familiarity of that item. Specifically, during the naming portion, participants were informed to type the name in Cantonese that corresponds to the concept represented by the image. Note that Cantonese characters include not only traditional Chinese characters, but also Cantonese unique characters (e.g., "冇" is a special character in Cantonese, which has the same meaning as "無" in traditional Chinese and "无" in simplified Chinese, and all three characters indicating "none" in English). Participants were encouraged to think and type the responses in Cantonese that could most accurately reflect the Cantonese naming of the picture, and they were also asked to offer precise responses and try not to use more than one word per concept. Vague, over-abstract, or obviously deviated responses of naming were told to be avoided. If they did not know the name of the picture, they should type "no" in the answer box. Moreover, consistent with the original MultiPic norming (Duñabeitia et al., 2022), participants needed to rate familiarity for each picture item on a continuous slider ranging from 1 (not familiar at all) to 100 (very familiar). Participants were encouraged to use the whole scale, rather than only the extreme numbers. The 100-point slider provided higher precisions than traditional 5-point or 7-point scales and participants were encouraged to choose more freely anywhere on the scale.

2.1.4 Analysis

Participants' responses were first cleaned with typos corrected, as well as eliminating any extraneous characters or digits. Additionally, trials of "no (i.e., do not know)" responses (3.20% of MultiPic data; 5.00% of McRae data) were excluded from further analysis. Furthermore, for an item, responses that only one participant said were defined as idiosyncratic responses and omitted from further analysis (9.47% of MultiPic data; 11.66% of McRae data).

After data cleaning and eliminating the "don't know" and idiosyncratic responses, we calculated the H statistics, modal response percentage, and concept agreement for each item in

MultiPic and McRae. H statistics is a value reflecting the function of items' alternative names, which could be calculated according to the formula below (Brodeur et al., 2014; Snodgrass & Vanderwart, 1980).

$$H = \sum_{i=1}^{k} P_i \log_2(1/P_i)$$

In this equation, k is the total number of names given to each image, excluding "no (i.e., don't know)" and idiosyncratic responses and P_i is the percentage of participants that one response was given. For a single picture item, a smaller H value would indicate a higher level of name agreement across participants (Brodeur et al., 2014).

For a particular picture, the modal response is the most frequent response, in other words, is provided by most participants. The corresponding English label from original norming for each picture is also provided. If other responses represented the same concept to the modal response, they were then treated as the same concept for concept agreement calculation (number of responses representing the same concept with the modal response/total number of responses). Additionally, if one picture item had two or more modal responses given simultaneously by the same number of participants, the other alternatives are provided in the column labeled "alternative modal". Additionally, familiarity for each picture item was calculated by averaging all participants who provided a familiarity rating on a scale of 1-100 for that picture. The mean and standard deviation of the familiarity rating were provided. Furthermore, since familiarity was a subjective rating score for each item from each participant, a two-way random-effect model intraclass correlation (ICC) was calculated. Based on Koo and Li (2016), a value higher than 0.75 indicates good reliability. Results showed that the familiarity rating across participants was highly reliable (MultiPic, 0.95; McRae, 0.92).

To quantify the visual complexity of all pictures, we used the *ImageDecomp* program developed by (Kardan et al., 2015) and (Berman et al., 2014). Images were decomposed into low-level features such as edge-related visual features, the shape of the histogram of pixel values, and color-related visual features. This program was then developed to capture the color and spatial properties of images. The measure of entropy (i.e., reflecting information content in an image) was used in the current study as the index of visual complexity. Compared to self-rated visual complexity, this approach provides an objective complexity measurement solely based on low-level visual features.

Furthermore, we obtained the frequency of the most dominant responses for all pictures from the Cantonese Web Corpus (CantoneseWaC) which consists of 42,669,08 tokens (up to 28th May 2023). The Sketch Engine database is made up of texts collected from the internet for more than 400 language corpora. The count, ratio (equals to the count divided by the total number of tokens), as well as the log10 transformed count are reported.

In summary, in the spreadsheet that contains final norming results for each dataset, for each picture item, we provide *language* (which is Cantonese), *picture code* (as an identifier to match with the original English norming), *picture group* (either in MultiPic250, MultiPic500 or McRae), *number of responses* (i.e., the total number of participants who rated the particular picture), *H statistics* (of name agreement), *English label* (match with the original English norming), *modal response* (response provided by most participants), *alternative modal* (response provided by the same number of participants as the modal response), *modal response percentage* (number of modal responses divided by the total number), *valid alternative response percentage* (number of valid alternative responses divided by the total number), *don't know response percentage* (number of times don't know responses divided by the total number), *idiosyncratic response percentage* (number of times N=1 responses divided by the total number), *familiarity (mean)*, *familiarity (standard deviation)*, *visual complexity*, *Sketchengine Cantonese frequency count*, *Sketchengine Cantonese frequency ratio*, *Sketchengine Cantonese frequency log10*, and *concept agreement percentage*. We then reported the descriptive statistics of these variables in the results section. All final norming results, as well as raw data, and analysis scripts are available online as psycholinguistic norms for Cantonese in the following link at Open Science Framework, https://osf.io/dz9j6/?view_only=a452d8a56c92430b9dedf21ac26b1bc1.

In addition to reporting the basic norm variables associated with each item, Pearson's correlational analyses were conducted to reveal relationships among several reported variables, including H statistics, modal response percentage, familiarity (mean), visual complexity, Sketchengine Cantonese frequency log10, and concept agreement. Furthermore, additional analyses were conducted with previously available norming in other languages both picture sets, to examine the similarity and variability across different languages. Last but not the least, we compared norming variables regarding their similarity and variability between the MultiPic and McRae databases, providing reference for the feasibility of combining the two picture sets.

2.2 Result

2.2.1 Descriptive statistics

In this section, we reported the descriptive statistics of picture properties including H statistics, modal response percentage, valid alternative response percentage, don't know response percentage, idiosyncratic response percentage, concept agreement percentage, familiarity (mean), visual complexity, as well as modal response word properties such as Sketchengine Cantonese frequency log10 (Table1). Pictures from two datasets are reported separately in parallel.

Table 1 Descriptive statistics of picture and word properties in MultiPic and McRae databases.

	MultiPic	McRae
Number of pictures	750	536
Number of responses	81	48
H statistics of name agreement	1.56 (0.86)	1.45 (0.81)
Modal response percentage	53.16 (23.59)	51.09 (24.32)
Valid alternative response percentage	34.19 (19.26)	32.25 (19.29)
Don't know response percentage	3.20 (6.58)	5.00 (8.94)
Idiosyncratic response percentage	9.46 (7.39)	11.67 (8.85)
Concept agreement percentage	78.85 (20.29)	73.81 (22.88)
Familiarity (mean)	75.97 (9.17)	74.89 (11.74)
Visual complexity	2.55 (0.87)	4.29 (1.94)
Log10 frequency	2.05 (0.91)	2.01 (0.91)

Values reported are means with standard deviations in the parentheses.

2.2.2 Correlational Analyses

As with other normative studies, correlational analyses were conducted to explore the relationships among different variables. Analyses were performed on the two sets separately and results are shown in Figure 1. Across two picture sets, the H statistics of name agreement, concept agreement, and familiarity rating were highly correlated with each other (ps < .001). Additionally, the familiarity was positively associated with target word frequency in MultiPic ($p < .001$), while the concept agreement was positively related to target word frequency in McRae ($p < .01$).

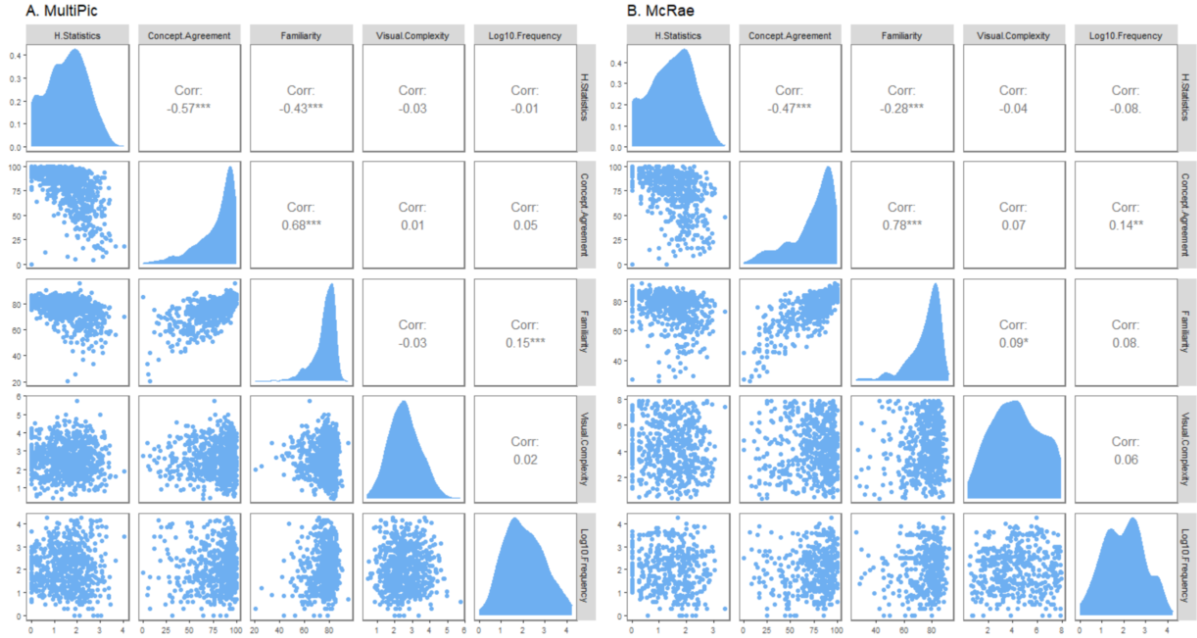

*Figure 1 The scatter and density plots of normed variables, and their correlations in A) MultiPic pictures, and B) McRae pictures. Reported correlations are r values. Significance levels are p < .05 *, p < .01 **, and p < .001***.*

2.2.3 Similarity and variability across languages

Because both the MultiPic and McRae sets were originally normed in English, and the MultiPic database was further normed in other languages, it provided us an opportunity to investigate language commonalities and differences (Table 2).

To compare the commonalities, we first performed a series of correlational analyses to compare normed variables across languages (r and *p* values can be found in Table 2). The current Cantonese norming and the original MultiPic English norming share common variables including the H statistics, as well as visual complexity (objective measurement in Cantonese norm vs. self-rating in English norm on a scale of 1-5). Correlational analyses showed that all these variables were highly comparable between English and Cantonese (*ps* < .001). Notably, a subset of 500 pictures in MultiPic is also available in Mandarin Chinese. Considering the similarities between Cantonese and Mandarin Chinese, we also compared the norms in these two languages on factors including the H statistics, and familiarity rating. Not surprisingly, all these variables were highly correlated with each other between Cantonese and Mandarin Chinese (*ps* < .001). The McRae concept norm provided variables including the self-rated familiarity of the concepts on a scale of 1-9, and the frequency of the modal name in KF

(Kučera et al., 1967) and BNC (British National Corpus, Leech, 1992). Correlational analyses showed that the familiarity rating was similar in Cantonese and English ($p < 0.001$). Besides, the modal name frequency in Cantonese was also comparable to that in English ($p < 0.001$).

In addition to the similarities, we also analyzed the differences across languages by running paired t-tests (t and *p* values can be found in Table 2). Comparing the norms available in the MultiPic whole set pictures, Cantonese (M = 1.55) received much higher H statistics (i.e., lower name agreement) than English (M = 0.77, $p < .001$). The objective and subjective visual complexity were not comparable because they were not on the same scales. In addition, Cantonese (M = 1.31) and Mandarin (M = 1.07) comparison on MultiPic500 pictures showed that Cantonese had significantly higher H statistics than Mandarin ($p < .001$), and native Cantonese speakers also reported being more familiar to the pictures than native Mandarin speakers ($p < .001$). Last but not least, although the frequency was not comparable because they were not on the same scales, the t test on the familiarity from the McRae set showed that the pictures in the Cantonese norming (M = 74.88) received significantly higher familiarity rating than the concepts in original rating on English concepts (M = 62.36, $p < .001$).

Table 2 Correlation and t-tests with MultiPic and McRae for the common indices across languages.

Current Cantonese	MultiPic Whole Set English		MultiPic500 Mandarin		McRae English	
	r	t	r	t	r	t
H statistics of name agreement	.46***	25.76***	.73***	9.06***	*NA*	*NA*
Familiarity	*NA*	*NA*	.91***	19.90***	.47***	14.65***
Visual complexity	.48***	*NA*	*NA*	*NA*	*NA*	*NA*
Log10 frequency	*NA*	*NA*	*NA*	*NA*	KF .16*** BNC .16***	*NA*

*NA indicates not available. Values reported are Pearson's r, and paired-t comparing with the same pictures from Cantonese norming. Significance levels are p < .001***.*

2.2.4 Comparison between MultiPic and McRae

The MultiPic and McRae picture databases shared 215 concepts in common. To demonstrate whether it is feasible to combine these two datasets, correlational analyses and t-tests were conducted to compare normed variables between the two datasets. Correlational analyses showed that the normed variables between the two picture sets were highly correlated with each other (H statistics, r = .74, $p < .001$; familiarity, r = .58, $p < .001$; visual complexity, r = .33, $p < .001$; concept agreement, r = .57, $p < .001$). Yet, t-tests results showed that the normed variables in the two picture sets were still significantly different from each other, with MultiPic pictures had higher H statistics (t = 1.98, $p = .04$), and McRae picture had higher familiarity rating (t = 2.53, $p = .01$), and higher visual complexity (t = 12.33, $p < .001$). The concept agreement between the two picture sets was not significantly different from each other (t = .87, $p = .38$).

2.3 Discussion

In the current experiment, we present Cantonese norming from typed responses of pictures and concepts from two databases: MultiPic and McRae. The MultiPic database contains 750 colored line drawings of a variety of concepts that have been normed in many languages. The McRae norm contains 536 concepts normed in English, and we sourced the corresponding pictures for these concepts from two picture databases (Brodeur et al., 2014; Moreno-Martínez & Montoro, 2012), as well as publicly available images from the internet (536 pictures in total). For each set, we reported normative variables including *H statistics*, *modal response*, *modal response percentage*, *valid alternative response percentage*, *don't know response percentage*, *idiosyncratic response percentage*, *familiarity*, *concept agreement*, *visual complexity*, *Sketchengine Cantonese frequency count, ratio* and *log10 transformation.*

We found that in each picture set, the H statistics, concept agreement, and familiarity were always significantly correlated with each other. Specifically, higher name agreement (i.e., lower H statistics) was associated with higher concept agreement and higher familiarity rating of that picture. These results are consistent with many previous studies. For example, in two previous norms in Mandarin Chinese norm (Chen & Zhu, 2015; Weekes et al., 2007), the name agreement and familiarity were highly correlated with each other. The current norming also brings a noteworthy contribution by introducing an objective approach to evaluate visual complexity. Traditionally, the assessment of image complexity has relied on individuals' self-rating using a specific scale. However, self-rating is inherently subjective, influenced by personal experiences and acquired knowledge (Donderi & McFadden, 2003). When it comes to image rating, Forsythe et al. (2008) found that subjective measures of visual complexity can be prone to bias, particularly due to familiarity effects.

In contrast, the current approach utilizes the *ImageDecomp* program, enhancing credibility by mitigating the influence of individual differences that would typically arise in subjective self-rating methods. We also validated our norming by comparing calculated variables in Cantonese to the norming in English and Mandarin Chinese on the same picture sets. We found that the common variables including H statistics, familiarity, visual complexity, and target word frequency between different languages in either MultiPic or McRae were all highly correlated with each other, indicating that picture name characteristics are similar across languages. These results nicely reflected the validity of the current study which extended picture norming of MultiPic and McRae concepts to Cantonese. Although the norming variables were in general correlated, they were still significantly different across languages, with the name agreement being lower but familiarity being higher in Cantonese than English and Mandarin. Although no previous cross-language comparison with Cantonese is available, studies comparing Mandarin Chinese and English have often reported that Mandarin norming had higher H statistics than English (Duñabeitia et al., 2022; Yoon et al., 2004). In comparison to Mandarin and English, Cantonese exhibits an even greater diversity of expressions for conveying the same concept. This could be due to the prevalent use of terms from various dialects to name identical objects, resulting in a broader range of linguistic variations, as exemplified by the participant's responses to a generic picture of a *nose*, such as *'鼻,' '鼻子,' '鼻公,' '鼻哥,'* and so on. The higher familiarity rating in Cantonese than Mandarin and English, might reflect the variations in self-reporting influenced by different cultural factors. The difference in familiarity ratings between English and Cantonese on the McRae set may also be potentially attributed to the fact that the English ratings were based solely on concepts, while the Cantonese ratings were conducted with pictures paired with the concept, making them imageable. Given these cross-language differences, it is important for future studies to incorporate these variables as control factors when making comparisons among different languages.

In addition to the cross-language comparison, we also compared the two picture sets on the common concepts. Results showed that all normed variables including name and concept agreement, picture familiarity, and visual complexity were all highly correlated between the two picture sets.

These results revealed a high level of comparability between the two sets across all variables. Thus, the merging of these datasets is plausible and has the potential to significantly expand the collection of available candidate pictures. While the norming variables showed correlation, and both sets have the commonality of being colored pictures, we acknowledge the stylistic distinctions between the two sets. By presenting these differences and similarities objectively, we aim to aid future researchers, especially those exploring the impacts of picture types, in making well-informed decisions about using these collections. Moreover, if researchers choose to combine the two sets, it is essential for them to recognize and consider these differences. We recommend that they include these normed variables as controls when selecting a combined set of picture stimuli.

Last but not least, one potential limitation in the current experiment is the absence of AoA data. Many studies commonly rely on subjective rating of AoA where participants estimate the age at which they learned a word (Chen & Zhu, 2015; Crepaldi et al., 2012; Momenian et al., 2021). Nevertheless, subjective reporting can be prone to various inaccuracies influenced by factors such as underestimation of vocabulary knowledge due to infantile amnesia, or a broader age range for respondents to choose from, and natural memory decay of early-acquired vocabulary (Xu et al., 2020). Alternatively, AoA information can be objectively obtained by analyzing the occurrence of a word in children's speech samples (Perret & Bonin, 2019). Unfortunately, unlike other languages, there are very few Cantonese-based speech corpus available. Despite attempting to retrieve AoA information from the MacArthur-Bates Communicative Development Inventories (Frank et al., 2017), our efforts were hindered by the unsatisfactory overlap between the available statistics and the items in the current experiment. Future norming studies on Cantonese are recommended to collect AoA data when feasible.

Overall, the present norming provides several linguistic variables in Cantonese that are important for selecting picture naming materials. It offers a timely and valuable alternative for researchers in the field of psycholinguistics, specifically those studying Cantonese production and lexical retrieval. To decide the efficacy of the current norming, it is crucial to further domenstrate how well these normed variables could predict picture naming reaction time. Therefore, we introduce the second experiment.

3. Experiment 2 Validating Typing Norming with Spoken Picture Naming

3.1 Methods

3.1.1 Participants

Thirty college students with comparable demographic profiles were recruited to participate the oral picture naming task (Mean age = 20.8 years, SD = 2.30 years, Range =18-25 years, 25 female). There were no repeated participants between the two experiments. All participants spoke Cantonese as their native and dominant language, and they were all multilinguals similar to the participants in Experiment 1. Among these participants, 79% participants reported their second language to be Mandarin Chinese, while 79% participants reported their third language to be English. The self-rated proficiency for Cantonese (Mean = 0.90, SD = 0.11) was comparable with the second language (Mean = 0.85, SD = 0.14; $p > .1$), which was significantly higher than the third language (Mean = 0.60, SD = 0.12; $p < .001$). Informed consent was obtained from each participant prior to the beginning of the study. The study protocols, procedures, and consent forms were approved by the Research Ethics Committee of University of Macau.

3.1.2 Stimuli and procedure

A subset of 300 pictures from MultiPic (170 pictures) and McRae (120 pictures) were selected for oral picture naming. During the picture naming task, pictures were presented and participants were instructed to overtly name the picture as quickly and accurately as possible. Each picture disappeared immediately after participants made a response or when the maximum response time of 3000 ms was reached. All pictures were pseudorandomly presented and participants had an opportunity to take a break in the half way. Before the formal experiment, ten pictures were presented to participants as practice. Stimuli were not repeated across the practice or the formal runs. The experiment was conducted using the E-Prime 3 software (Psychology Software Tools, Pittsburgh, PA). Participants' responses were recorded using a microphone and a digital recorder.

3.1.3 Data analysis

Participants' responses were coded for accuracy (ACC) and reaction times (RT) based on the recordings. Responses were marked as incorrect (7.08% of MultiPic data; 9.47% of McRae data) if they were omissions (including no responses, or do not know), dysfluencies (including hesitation, repetition, or correction). Idiosyncratic responses were also identified and excluded from further analysis (6.73% of MultiPic data; 7.89% of McRae data). Reaction times (RTs) were calculated using customized PRAAT scripts (Boersma & Weenink, 2002), which identified speech onsets by searching the recordings for pitch deviations within the audio recording. The onsets were then manually verified by using the audio and visual speech stream. After then, RTs were calculated as the difference between picture onsets (from E-Prime output) and the response onsets. Only RTs for correct and non-idiosyncratic trials, longer than 200 ms, and within 2.5 SDs were included in further analyses (11.24 % removed). In addition to ACC and RT, other normed variables reported from Experiment 1 (i.e., typing normed variables) were also calculated based on data from participants in Experiment 2 (i.e., spoken normed variables).

To investigate whether the typing and spoken normed variables could predict spoken reaction times, generalized linear mixed-effect regression analyses were conducted, employing *lmer* function in the *lme4* package, respectively (Bates et al., 2014) in the R environment (R Core Team, 2014). This approach allows random effects therefore has the advantage of considering individual data points and controls for variation across participants and items simultaneously, producing more generalizable results. When defining the models, we started from the full model as suggested in previous literature (Barr et al., 2013). If the full models do not converge, then we took out item level random intercept but keeping the random slopes. For either spoken normed variables, or typing normed variables, the regression included the trial-level reaction times as the dependent variable, and fixed effects for across-participant item-level H statistic of name agreement, visual complexity, and target word log frequency as independent variables. Additionally, for both regressions, random slopes of H statistics, visual complexity and target word log frequency were included, allowing for variability in the impact of these predictors across items. Furthermore, the familiarity rating was included as one independent variable in regression, and the values from trial level were utilized for the spoken norming regression. Yet only values from across-participant item-level familiarity rating were used for the typing norming regression because participants were different from spoken and typing experiments, but the random slope of familiarity rating was then included in the regression. Note that the concept agreement was not included because of the collinearity concern with H statistics (*r* = -. 47). Since all participants were multilinguals, additional regression analyses were also conducted to investigate the relationships between language experience (i.e., proficiency and number of languages) and picture naming reaction time.

In addition, to further investigate the consistency between different ways of testing—typing in Experiment 1 and speaking in Experiment 2, correlational analyses were conducted between the two experiments on the H statistics of name agreement and familiarity rating.

3.2 Results

First, a mixed-effect regression was conducted on RT, including four independent variables from the spoken norming from the current experiment (i.e., familiarity, H statistics, visual complexity, and target word log frequency). The model can be specified as: $RT\ (trial\ level) = \beta_0 + \beta_1\ familiarity\ (trial\ level) + \beta_2\ H\ statistics + \beta_3\ visual\ complexity + \beta_4\ Frequency + \mu_2\ H\ statistics + \mu_3\ visual\ complexity + \mu_4\ Frequency + e$. Results showed that faster reaction time was significantly predicted by lower H statistics of name agreement (Figure 2A, β = .09, SE = .01, t = 6.64, $p < .001$), higher familiarity rating (Figure 2B, β = -.02, SE = .005, t = -3.19, $p = .001$), higher target word frequency (Figure 2C, β = -.03, SE = .01, t = -2.44, $p = .02$). Yet, no significant relationship was found between reaction time and visual complexity (β = .02, SE = .01, t = 1.71, $p = .09$).

Similarly, another mixed-effect regression was conducted on RT, including the same four independent variables from the original typing norming from Experiment 1. *The model can be specified as:* $RT\ (trial\ level) = \beta_0 + \beta_1\ familiarity\ (across\text{-}participant\ item\text{-}level) + \beta_2\ H\ statistics + \beta_3\ visual\ complexity + \beta_4\ Frequency + \mu_0 + \mu_1\ familiarity\ (across\text{-}participant\ item\text{-}level) + \mu_2\ H\ statistics + \mu_3\ visual\ complexity + \mu_4\ Frequency + e$. Results showed that faster reaction time was significantly predicted by lower H statistics of name agreement (Figure 2A', β = .06, SE = .01, t = 5.46, $p < .001$), higher familiarity rating (Figure 2B', β = -.15, SE = .01, t = -11.92, $p < .001$), higher target word frequency (values same to Figure 2C, β = -.05, SE = .01, t = -5.01, $p < .001$). Yet, no significant relationship was found between reaction time and visual complexity (β = .02, SE = .01, t = 1.40, $p = .17$).

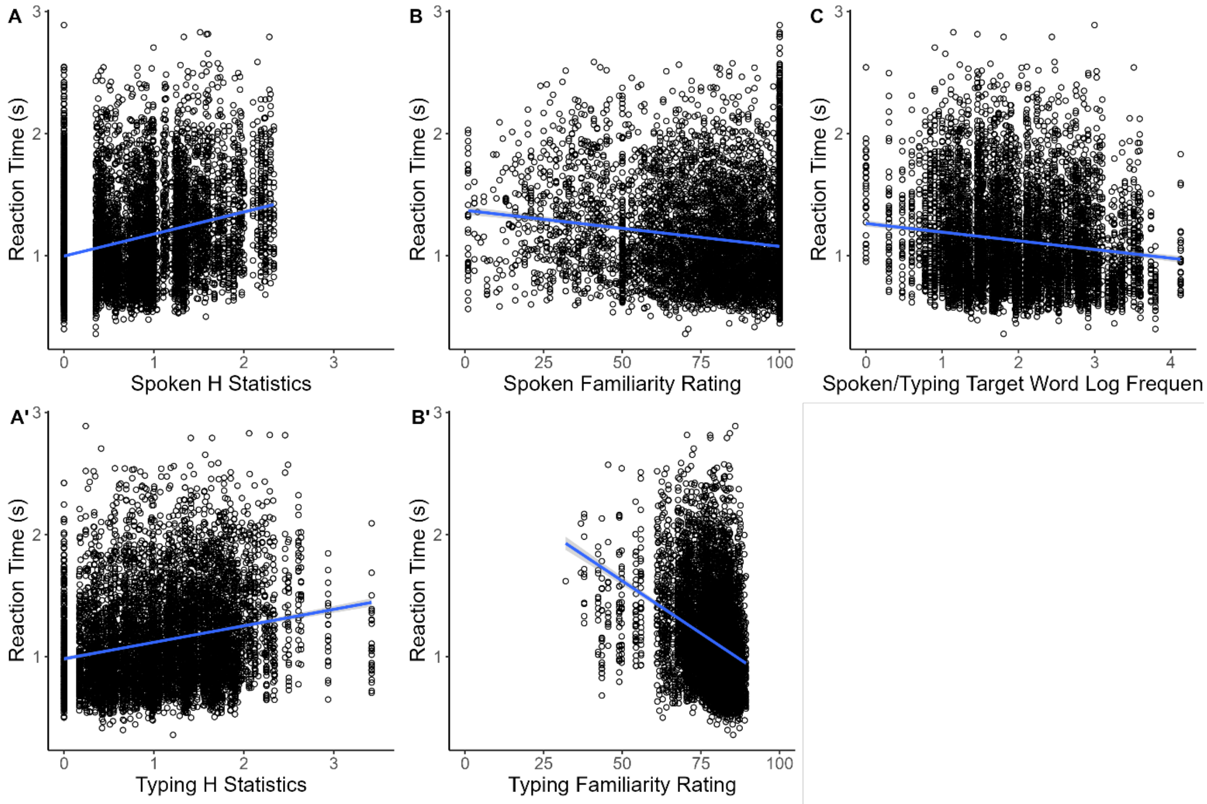

Figure 2. The predictivity effect of spoken (top panel) and typing (bottom panel) variables to spoken reaction times. Across both modules, higher name agreement (i.e., lower H statistics, A and A'), higher familiarity rating (B based on trial level rating, and B' based on across-participant item level rating), and higher target word frequency for spoken/typing (C) were significantly associated with faster reaction time.

Exploration on the relationship between language proficiency and reaction time showed that higher L1 proficiency was significantly associated with shorter reaction time (Figure 3A, β = -.13, SE = .04, $p < .001$), while higher L2 (Figure 3B, β = .32, SE = .03, $p < .001$) and L3 (Figure 3C, β = .12, SE = .04, $p = .004$) proficiency was significantly associated with longer reaction time. Yet, the number of languages spoken did not significantly affect picture naming reaction times (Figure 3D), F (1, 28) = .59, $p = .45$.

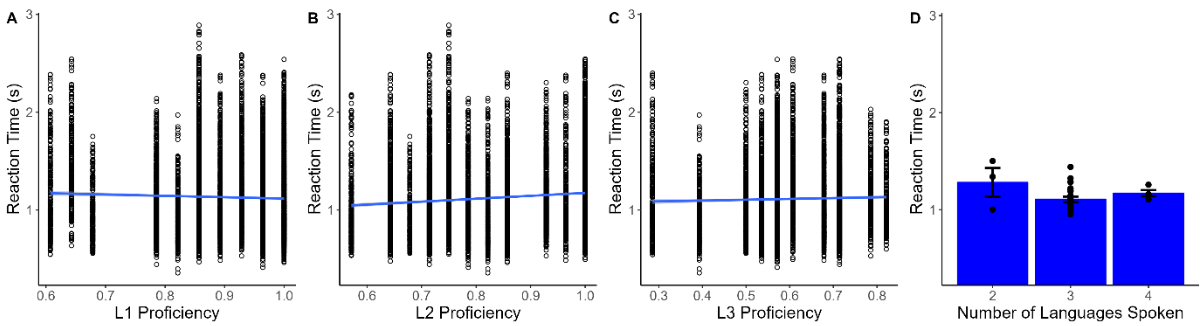

Figure 3. The relationship between self-rated language proficiency and picture naming reaction times. Higher native language proficiency (A), lower second (B) and third language (C) proficiency were associated with

faster reaction times in spoken picture naming. Black dots in (A) (B) (C) represent trial level RT. Yet, there was no effect of number of languages on reaction times (D). Black dots in D represent subject level mean RT.

To compare the consistency between different modules of testing, correlation analyses were conducted on the normed variables calculated from each experiment on the subset stimuli in Experiment 2 (i.e., 300 pictures, Figure 4). Results showed that both the H statistics of name agreement (r = .59, $p < .001$), and the frequency rating (r = .87, $p < .001$) were significantly correlated between the two experiments, demonstrating a high consistency between typing and spoken responses.

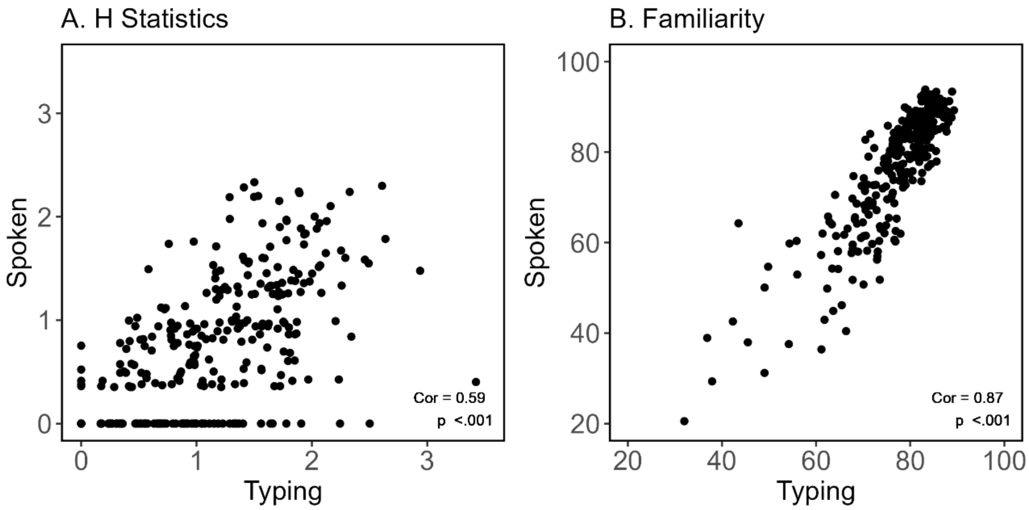

Figure 4 Consistency on H statistics (A) and familiarity rating (B) calculated from typing and spoken responses.

3.3 Discussion

In the current experiment, we validated the predictive capacity of the normed variables generated from typing to spoken picture naming. A different group of participants with comparable demographic characteristics named a subset of the original pictures orally and corresponding normed variables were calculated from their spoken responses.

Correlational analyses on H statistics and familiarity rating suggested that different modules of norming (i.e., typing vs. spoken) demonstrated a high level of similarity. Despite these high similarities across modules, it is important to highlight a distinct pattern observed in the typing module, where a greater variety of responses was present compared to the spoken module (i.e., H statistics from typing displayed fewer zero values and a wider range than those from speaking). This discrepancy may be related to participants having the opportunity to take more time to formulate and type their responses, thereby introducing larger variations in word selection during typing. Essentially, spoken responses may engage in more automatic processes, whereas typing responses may reflect a more deliberate and controlled cognitive effort. Besides the comparability on normed variables between the two modules, typing and spoken norming variables stably predicted reaction times in picture naming in a similar manner. Specifically, higher name agreement, familiarity, and target word frequency are all associated with slower responses, indicating more difficult retrieval processes during production. These results are consistent with previous studies normed in other languages (Alario et al., 2004; Perret & Bonin, 2019), including Mandarin Chinese and Cantonese (Chen & Zhu, 2015; Momenian et al., 2021; Weekes et al., 2007). Additionally, these factors are assumed to exert their influence at different levels of picture naming (Alario et al., 2004; Graves et al., 2007). Specifically, picture familiarity reflects how frequently the presented object in an image is encountered, thereby impacting the processing of object and concept identification. Additionally, name agreement indicates the level of consensus among participants in labeling a specific picture. Lower agreement suggests more lexical labels competing for selection. Therefore, when examining picture naming, name agreement should be one of the most important predictors to consider. Last but not least, the current study, along with numerous previous studies, highlights the effect of word frequency on picture naming reaction times, by modulating the difficulty of lexical selection. The effect of word frequency on picture naming has been consistently observed and recommended as a significant predictor of picture naming performance (Jescheniak & Levelt, 1994; Oldfield & Wingfield, 1965).

Although the current sample were all native Cantonese speakers, they were all multilinguals who spoke at least two languages. Although the number of languages spoken did not significantly affect picture naming reaction times, we found that higher Cantonese proficiency was associated with faster Cantonese naming. Yet, higher proficiency in other non-native languages was associated with slower naming in the native Cantonese, consistent with many previous studies and the Frequency Lag Hypothesis (Gollan et al., 2008; Gollan et al., 2011). It should be noted that although the current sample well represents the local population and even all Cantonese speaking population, it might not well represent Cantonese monolingual speakers. Researchers who are interested in Cantonese monolingual population should utilize this norm with caution.

4. Conclusion

In the current study, we normed a total of 1286 pictures from multiple resources in Cantonese. Normed variables including *name agreement, concept agreement, familiarity, visual complexity,* and *target word frequency* are provided. We compared the current Cantonese norming to other languages, demonstrating similarities and variations among different languages. We also embraced the multilingual nature of the current population and found that higher Cantonese proficiency but lower non-native language proficiency was associated with better spoken picture naming. Last but not least, we validated the predictivity power of normed variables calculated from typing responses to spoken picture naming performance and the consistency between typing and spoken responses. All raw data, analysis scripts, and final norming results are available online as psycholinguistic norms for Cantonese in the following link at Open Science Framework, https://osf.io/dz9j6/?view_only=a452d8a56c92430b9dedf21ac26b1bc1.

Overall, the present norming provides a timely and valuable alternative for researchers in the field of psycholinguistics, specifically those studying Cantonese production and lexical retrieval. It offers a reliable method for selecting Cantonese picture naming materials, proving to be highly beneficial for researchers seeking accurate and relevant resources.

Declarations

Funding

This project was funded by the National Natural Science Foundation of China (32200845), the Science and Technology Development Fund, Macao S.A.R (FDCT, 0153/2022/A), and the Start-up Research Grant (SRG2022-00003-ICI) and Multi-Year Research Grant (MYRG2022-00148-ICI) from the University of Macau to Haoyun Zhang.

Conflicts of interest

Authors declare that they have no competing interests.

Ethics approval

The study protocols, procedures, and consent forms were approved by the Research Ethics Committee of University of Macau.

Consent to participate

Informed consent was obtained from each participant prior to the beginning of the study.

Consent for publication

No identifiable information related to participants is reported in the article.

Availability of data and materials

All raw data, analysis scripts, and final norming results are available online as psycholinguistic norms for Cantonese in the following link at Open Science Framework, https://osf.io/dz9j6/?view_only=a452d8a56c92430b9dedf21ac26b1bc1.

The raw data and the analysis scripts are available upon request.

Acknowledgments

This project was funded by the National Natural Science Foundation of China (32200845), the Science and Technology Development Fund, Macao S.A.R (FDCT, 0153/2022/A), and the Start-up Research Grant (SRG2022-00003-ICI) and Multi-Year Research Grant (MYRG2022-00148-ICI) from the University of Macau to Haoyun Zhang. We thank the staff at the Centre for Cognitive and Brain Sciences (CCBS) at the University of Macau where the study was conducted, for their support. Authors declare that they have no competing interests. We thank Dr. Michele Diaz and members in the Language and Aging Lab at Penn State University for their help with sourcing pictures for the McRae concepts. We would like to acknowledge the McRae concept set (McRae et al., 2005), as well as the BOSS and Moreno picture norms (Brodeur et al., 2014; Francisco J. Moreno-Martínez & Montoro, 2012).

Reference

Adlington, R. L., Laws, K. R., & Gale, T. M. (2009). The Hatfield Image Test (HIT): A new picture test and norms for experimental and clinical use. *Journal of clinical and experimental neuropsychology*, *31*(6), 731-753

Alario, F.-X., Ferrand, L., Laganaro, M., New, B., Frauenfelder, U. H., & Segui, J. (2004). Predictors of picture naming speed. *Behavior Research Methods, Instruments, & Computers*, *36*, 140-155.

Bangalore, S., Robson, H., & Astell, A. J. (2022). Standardizing norms for 180 coloured Snodgrass and Vanderwart pictures in Kannada language. *PLoS ONE*, *17*(4). https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0266359

Barbarotto, R., Laiacona, M., Macchi, V., & Capitani, E. (2002). Picture reality decision, semantic categories and gender: A new set of pictures, with norms and an experimental study. *Neuropsychologia*, *40*(10), 1637-1653

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255-278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

Berman, M. G., Hout, M. C., Kardan, O., Hunter, M. R., Yourganov, G., Henderson, J. M., Hanayik, T., Karimi, H., & Jonides, J. (2014). The perception of naturalness correlates with low-level visual features of environmental scenes. *PLoS ONE*, *9*(12).

Boersma, P., & Weenink, D. (2002). Praat 4.0: a system for doing phonetics with the computer [Computer software]. *Amsterdam: Universiteit van Amsterdam*.

Bonin, P., Peereman, R., Malardier, N., Méot, A., & Chalard, M. (2003). A new set of 299 pictures for psycholinguistic studies: French norms for name agreement, image agreement, conceptual familiarity, visual complexity, image variability, age of acquisition, and naming latencies. *Behavior Research Methods, Instruments, & Computers*, *35*, 158-167

Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The Bank of Standardized Stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. *PLoS ONE*, *5*(5).

Brodeur, M. B., Guérard, K., & Bouras, M. (2014). Bank of standardized stimuli (BOSS) phase II: 930 new normative photos. *PloS one*, *9*(9), e106953.

Brysbaert, M., Stevens, M., De Deyne, S., Voorspoels, W., & Storms, G. (2014). Norms of age of acquisition and concreteness for 30,000 Dutch words. *Acta Psychologica*, *150*, 80-84. https://www.sciencedirect.com/science/article/pii/S0001691814000985

Census, S. D. H. K. (2020). Use of Language in Hong Kong 香港的語言使用情況.

Chen, Y., & Zhu, L. (2015). Predictors of Action Picture Naming in Mandarin Chinese. *Acta Psychologica Sinica*, *47*(1), 11

Crepaldi, D., Che, W.-C., Su, I.-F., & Luzzatti, C. (2012). Lexical-Semantic Variables Affecting Picture and Word Naming in Chinese: A Mixed Logit Model Study in Aphasia. *Behavioural Neurology*, *25*(3), 165-184. https://doi.org/10.3233/ben-2012-119002

Cuetos, F., & Alija, M. (2003). Normative data and naming times for action pictures. *Behavior Research Methods, Instruments, & Computers*, *35*(1), 168-177

Cuetos, F., Ellis, A. W., & Alvarez, B. (1999). Naming times for the Snodgrass and Vanderwart pictures in Spanish. *Behavior Research Methods, Instruments, & Computers*, *31*(4), 650-658. https://doi.org/10.3758/BF03200741

Dan-Glauser, E. S., & Scherer, K. R. (2011). The Geneva affective picture database (GAPED): a new 730-picture database focusing on valence and normative significance. *Behavior Research Methods*, *43*, 468-477.

De Winter, J., & Wagemans, J. (2004). Contour-based object identification and segmentation: Stimuli, norms and data, and software tools. *Behavior Research Methods, Instruments, & Computers*, *36*(4), 604-624

Dell'acqua, R., Lotto, L., & Job, R. (2000). Naming times and standardized norms for the italian PD/DPSS set of 266 pictures: Direct comparisons with American, English, French, and Spanish published databases. *Behavior Research Methods, Instruments, & Computers*, *32*(4), 588-615. https://doi.org/10.3758/BF03200832

Donderi, D. C., & McFadden, S. (2003). A single marine overlay display is more efficient than separate chart and radar displays. *Displays*, *24*(4-5), 147-155

Duñabeitia, J. A., Baciero, A., Antoniou, K., Antoniou, M., Ataman, E., Baus, C., Ben-Shachar, M., Çağlar, O. C., Chromý, J., Comesaña, M., Filip, M., Đurđević, D. F., Dowens, M. G., Hatzidaki, A., Januška, J., Jusoh, Z., Kanj, R., Kim, S. Y., Kırkıcı, B., . . . Pliatsikas, C. (2022). The Multilingual Picture Database. *Scientific Data*, *9*(1), 431

Duñabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Pliatsikas, C., Smolka, E., & Brysbaert, M. (2018). MultiPic: A standardized set of 750 drawings with norms for six European languages. *Quarterly Journal of Experimental Psychology*, *71*(4), 808-816

Ellis, A. W., & Morrison, C. M. (1998). Real age-of-acquisition effects in lexical retrieval. *Journal of experimental psychology: learning, Memory, and cognition*, *24*(2), 1939-1285.

Fah, E. L. E. (2017). Profile of the Singapore Chinese dialect groups. *Statistics*.

Forsythe, A., Mulhern, G., & Sawey, M. (2008). Confounds in pictorial sets: The role of complexity and familiarity in basic-level picture processing. *Behavior Research Methods*, *40*(1), 116-129.

Francisco J. Moreno-Martínez, & Montoro, P. R. (2012). An ecological alternative to Snodgrass & Vanderwart: 360 high quality colour images with norms for seven psycholinguistic variables. *PLoS ONE*, *7*(5).

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: an open repository for developmental vocabulary data. *Journal of Child Language*, *44*(3).

Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language*, *58*(3), 787-814. https://www.sciencedirect.com/science/article/pii/S0749596X07000836

Gollan, T. H., Slattery, T. J., Goldenberg, D., Van Assche, E., Duyck, W., & Rayner, K. (2011). Frequency drives lexical access in reading but not in speaking: The frequency-lag hypothesis. *Journal of Experimental Psychology: General*, *140*(2), 186-209.

Graves, W. W., Grabowski, T. J., Mehta, S., & Gordon, J. K. (2007). A Neural Signature of Phonological Access: Distinguishing the Effects of Word Frequency from Familiarity and Length in Overt Picture Naming. *Journal of Cognitive Neuroscience*, *19*(4), 617-631.

Janssen, N., Pajtas, P. E., & Caramazza, A. (2011). A set of 150 pictures with morphologically complex English compound names: Norms for name agreement, familiarity, image agreement, and visual complexity. *Behavior Research Methods*, *43*, 478-490.

Jescheniak, J. D., & Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of experimental psychology: learning, Memory, and cognition*, *20*(4), 824-843. https://doi.org/10.1037/0278-7393.20.4.824

Kardan, O., Demiralp, E., Hout, M. C., Hunter, M. R., Karimi, H., Hanayik, T., Yourganov, G., Jonides, J., & Berman, M. G. (2015). Is the preference of natural versus man-made scenes driven by bottom–up processing of the visual features of nature? *Frontiers in Psychology*, *6*, 1664-1078.

Kim, H., & Na, D. L. (1999). BRIEF REPORT normative data on the Korean version of the Boston naming test. *Journal of clinical and experimental neuropsychology*, *21*(1), 127-133

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155-163.

Kremin, H., Akhutina, T., Basso, A., Davidoff, J., De Wilde, M., Kitzing, P., Lorenz, A., Perrier, D., van der Sandt-Koenderman, M., & Vendrell, J. (2003). A cross-linguistic data bank for oral picture naming in Dutch, English, German, French, Italian, Russian, Spanish, and Swedish (PEDOI). *Brain and Cognition*, *53*(2), 243-246

Kremin, H., Perrier, D., Wilde, M. D., Dordain, M., Bayon, A. L., Gatignol, P., Rabine, C., Corbineau, M., Lehoux, E., & Arabia, C. (2001). Factors predicting success in picture naming in Alzheimer's disease and primary progressive aphasia. *Brain and Cognition*, *46*(1), 180-183. https://www.sciencedirect.com/science/article/pii/S0278262600912703

Kučera, H., Francis, W., Twaddell, W. F., Marckworth, M. L., Bell, L. M., & Carroll, J. B. (1967). Computational analysis of present-day American English. *International Journal of American Linguistics*, *35*(1), 71-75.

LaGrone, S., & Spieler, D. H. (2006). Lexical competition and phonological encoding in young and older speakers. *Psychology and Aging*, *21*(4), 804-809.

Leech, G. (1992). 100 million words of English: the British National Corpus (BNC). *Language research*, *28*(1), 1-13.
Li, D. C. S. (2006). Chinese as a lingua franca in greater China. *Annual Review of Applied Linguistics*, *26*, 149-176
Liu, Y., Hao, M., Li, P., & Shu, H. (2011). Timed Picture Naming Norms for Mandarin Chinese. *PLoS ONE*, *6*(1).
Magnié, M. N., Besson, M., Poncet, M., & Dolisi, C. (2003). The Snodgrass and Vanderwart set revisited: Norms for object manipulability and for pictorial ambiguity of objects, chimeric objects, and nonobjects. *Journal of clinical and experimental neuropsychology*, *25*(4), 521-560
McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods*, *37*(4), 547-559
Migo, E. M., Montaldi, D., & Mayes, A. R. (2013). A visual object stimulus database with standardized similarity information. *Behavior Research Methods*, *45*, 344-354.
Momenian, M., Bakhtiar, M., Chan, Y. K., Cheung, S. L., & Weekes, B. S. (2021). Picture naming in bilingual and monolingual Chinese speakers: Capturing similarity and variability. *Behavior Research Methods*, *53*(4), 1677-1688
Momenian, M., Lau, K.-Y. D., & Bakhtiar, M. (2022). Developing psycholinguistic norms for action pictures in Cantonese. *Applied Neuropsychology: Adult*, *0*(0), 1-8
Moreno-Martínez, F. J., & Montoro, P. R. (2012). An ecological alternative to Snodgrass & Vanderwart: 360 high quality colour images with norms for seven psycholinguistic variables. *PloS One*, *7*(5), e37527.
Newman, R. S., & German, D. J. (2005). Life Span Effects of Lexical Factors on Oral Naming. *Language and Speech*, *48*(2), 123-156. https://doi.org/10.1177/00238309050480020101
Ni, L., Liu, Y., Yu, W., & Fu, X. (2019). The China Image Set (CIS): A New Set of 551 Colored Photos With Chinese Norms for 12 Psycholinguistic Variables. *Frontiers in Psychology*, *10*
Nishimoto, T., Miyawaki, K., Ueda, T., Une, Y., & Takahashi, M. (2005). Japanese normative set of 359 pictures. *Behavior Research Methods*, *37*(3), 398-416
Nishimoto, T., Ueda, T., Miyawaki, K., Une, Y., & Takahashi, M. (2010). A normative set of 98 pairs of nonsensical pictures (droodles). *Behavior Research Methods*, *42*, 685-691
Oldfield, R. C., & Wingfield, A. (1965). Response Latencies in Naming Objects. *Quarterly 4 Journal of Experimental Psychology*, *17*(4), 273-281.275. https://doi.org/10.1080/17470216508416445
Perret, C., & Bonin, P. (2019). Which variables should be controlled for to investigate picture naming in adults? A Bayesian meta-analysis. *Behavior Research Methods*, *51*(6), 2533-2545
Pind, J., Jónsdóttir, H., Gissurardóttir, H. s., & Jónsson, F. (2000). Icelandic Norms for the Snodgrass and Vanderwart (1980) Pictures: Name and Image Agreement, Familiarity, and Aage of Acquisition [2000 The Scandinavian Psychological Associations]. *Scandinavian Journal of Psychology*, *41*(1), 41-48. https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-9450.00169
Pind, J., & Tryggvadóttir, H. B. (2002). Determinants of picture naming times in Icelandic [2002 The Scandinavian Psychological Associations]. *Scandinavian Journal of Psychology*, *43*(3), 221-226. https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-9450.00290
R Core Team. (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. *ISBN 3-900051-07-0*.
Rodríguez-Ferreiro, J., Davies, R., González-Nosti, M., Barbón, A., & Cuetos, F. (2009). Name agreement, frequency and age of acquisition, but not grammatical class, affect object and action naming in Spanish speaking participants with Alzheimer's disease. *Journal of Neurolinguistics*, *22*(1), 37-54. https://www.sciencedirect.com/science/article/pii/S0911604408000493
Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. *Perception*, *33*(2), 217-236
Salmon, J. P., McMullen, P. A., & Filliter, J. H. (2010). Norms for two types of manipulability (graspability and functional usage), familiarity, and age of acquisition for 320 photographs of objects. *Behavior Research Methods*, *42*, 82-95

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(2), 174-215.

Snow, D. (2008). Cantonese as written standard? *Journal of Asian Pacific Communication*, *18*(2), 190-208

Tsaparina, D., Bonin, P., & Méot, A. (2011). Russian norms for name agreement, image agreement for the colorized version of the Snodgrass and Vanderwart pictures and age of acquisition, conceptual familiarity, and imageability scores for modal object names. *Behavior Research Methods*, *43*(4), 1085-1099. https://doi.org/10.3758/s13428-011-0121-9

Valente, A., Bürki, A., & Laganaro, M. (2014). ERP correlates of word production predictors in picture naming: a trial by trial multiple regression analysis from stimulus onset to response. *Frontiers in Neuroscience*, *8*. https://www.frontiersin.org/articles/10.3389/fnins.2014.00390

Viggiano, M. P., Vannucci, M., & Righi, S. (2004). A new standardized set of ecological pictures for experimental and clinical research on visual object processing. *Cortex*, *40*(3), 491-509

Vollmann, R., & Wooi, S. (2018). Chinese identities in multilingual Malaysia. *89*, 35-61.

Wang, R., Huang, S., Zhou, Y., & Cai, Z. G. (2020). Chinese character handwriting: A large-scale behavioral study and a database. *Behavior Research Methods*, *52*(1), 82-96

Weekes, B. S., Shu, H., Hao, M., Liu, Y., & Tan, L. H. (2007). Predictors of timed picture naming in Chinese. *Behavior Research Methods*, *39*(2), 335-342

Xu, X., Li, J., & Guo, S. (2020). Age of acquisition ratings for 19,716 simplified Chinese words. *Behavior Research Methods*, *53*, 558-573.

Yoon, C., Feinberg, F., Luo, T., Hedden, T., Gutchess, A. H., Chen, H.-Y. M., Mikels, J. A., Jiao, S., & Park, D. C. (2004). A cross-culturally standardized set of pictures for younger and older adults: American and Chinese norms for name agreement, concept agreement, and familiarity. *Behavior Research Methods, Instruments, & Computers*, *36*(4), 639-649.

Yu, H. (2022). Cantonese migrant networks, white supremacy, and the political utility of apologies in Canada. *Journal of Ethnic and Migration Studies*, *48*(4), 969-988

Zannino, G. D., Perri, R., Salamone, G., Di Lorenzo, C., Caltagirone, C., & Carlesimo, G. A. (2010). Manipulating color and other visual information influences picture naming at different levels of processing: Evidence from Alzheimer subjects and normal controls. *Neuropsychologia*, *48*(9), 2571-2578. https://www.sciencedirect.com/science/article/pii/S002839321000182X

Zhou, D., & Chen, Q. (2017). Color Image Norms in Mandarin Chinese. *Frontiers in Psychology*, *8*