

Answering tendencies on questionnaires: Comparing Mandarin and Bahasa Malaysia versus English

International Journal of Bilingualism

1–16

© The Author(s) 2024



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/13670069241246523

journals.sagepub.com/home/ijb**Zhong Jian Chee**

University of Nottingham Malaysia, Malaysia; University of Aberdeen, UK

Yee Thung Lee 

University of Nottingham Malaysia, Malaysia; University of Leeds, UK

Omid R. Fani and Tze Wei Yong

University of Nottingham Malaysia, Malaysia

Marieke de Vries 

University of Nottingham Malaysia, Malaysia; Utrecht University, The Netherlands

Abstract

Aims Conducting psychological research in different countries and cultures necessitates measures in different languages. However, the language of a measure might influence responses, even within the same multilingual individual. The cultural accommodation theory proposes that one's association with a language influences their responses. Moreover, response styles (RSs), such as an extreme or acquiescence RS, might systematically affect responses regardless of the content of the measure. These effects were reported on culture-related measures but are unclear on culture-free measures. **Methodology and analyses** We aimed to investigate the effects of language on psychological measures that do not explicitly examine cultural factors. Multilingual Malaysians ($n = 111$) filled in the Adult Executive Functioning Inventory (ADEXI), the Depression, Anxiety, and Stress Scale—21 items (DASS-21), the Brief Coping Orientation to Problems Experienced Inventory (Brief COPE), the Pediatric Quality of Life Inventory (PedsQL), and the Traditional Masculinity-Femininity Scale (TMF) in Bahasa Malaysia and English, or in Mandarin and English. **Findings** There were no language differences on the ADEXI and TMF. However, several subscales of the Brief COPE, the Stress subscale of the DASS, and the PedsQL scores were higher in Mandarin than in English. On the Brief COPE and the PedsQL, there were also differences in RS between Mandarin and English, which might explain (part of) these differences. There were no differences between Bahasa Malaysia and English in scores. However, there was a more extreme

Corresponding author:

Marieke de Vries, Development & Education of Youth in Diverse Societies (DEEDS), Faculty of Social and Behavioural Sciences, Utrecht University, Martinus J. Langeveldgebouw, room F2.14, 3584 CS Heidelberglaan 1, Utrecht, The Netherlands.

Email: m.devries7@uu.nl

RS in English than in Bahasa Malaysia and a more acquiescence RS in Bahasa Malaysia than in English on the Brief COPE. These differences suggest that the measures are not culture-free or that previously reported language differences did not result from culture alone. The language of a measure might be an additional important factor. When using different translations of the same measure, it is important to take cultural accommodation and RS into account.

Keywords

Language, Mandarin, Bahasa Malaysia, questionnaires, multilingual

The current globalization and the growing number of people with multi-cultural backgrounds warrant a need to evaluate assessment tools in different countries and different languages. It is important to test the cross-cultural, or cross-language validity of questionnaires. Past studies have shown that language might influence scores, even within the same person. Bilingual participants can produce different responses according to the language of the instrument (Chee & de Vries, 2021; Peytcheva, 2018). This suggests that different languages can elicit different responses, and hence scores among participants from different countries or backgrounds (Richard & Toffoli, 2009). These differences in answers do not reflect a true difference in questionnaire content but reflect the cultural background of the participant or language interpretation. Different hypotheses try to explain this effect. The cultural accommodation hypothesis (Chen et al., 2014) and the ethnic affirmation hypothesis (Bond & Yang, 1982; K. S. Yang & Bond, 1980) have been proposed to explain the different answering tendencies in different languages. Moreover, response style (RS) might be influenced by language (Harzing, 2006).

Cultural accommodation takes place when participants provide responses that conform to the cultural system they associate with the language of the instrument (Chen et al., 2014). For instance, Chinese–English bilinguals demonstrate more dialectical thinking and variability in their personality and behavior on Chinese than on English written and verbal assessments (Chen et al., 2014). The written assessments involved ratings on the Dialectical Self Scale, Sino-American Person Perception Scale, and Big Five Inventory, whereas the verbal assessments involved conversing with a Caucasian and a Chinese interviewer in English and Cantonese. Similarly, among Chinese–English bilinguals, Western values such as individualism are accentuated when participants answer in English, while Chinese cultural values such as collectivism are accentuated when they answer in Chinese, their first language (Kemmelmeyer & Cheng, 2004; Richard & Toffoli, 2009).

In contrast, the ethnic affirmation hypothesis suggests that participants would provide responses that align with their ethnic culture when answering questions in their non-ethnic cultural language (Chen & Bond, 2007; K. S. Yang & Bond, 1980). For instance, Mandarin–English bilinguals were shown to demonstrate more identification with Chinese culture when answering in English than in Mandarin (Bond & Yang, 1982). It is suggested that the manifestation of cultural accommodation or ethnic affirmation depends on the importance of content values in a questionnaire to the participants (Bond & Yang, 1982). Ethnic affirmation tends to take place when the questions tap on cultural values or beliefs important to the participants, while cultural accommodation tends to occur when the content values are less important (Richard & Toffoli, 2009). Importantly, ethnic affirmation has not been replicated as much as cultural accommodation.

RS is a participant's tendency to provide a systematic response regardless of what the questionnaire intends to measure (Baumgartner & Steenkamp, 2001). Two types of RS are commonly observed; the tendency to provide extreme (Extreme RS) or middle responses (Middle RS) on a

Likert-type Scale, and the tendency to Agree (acquiescence RS) or Disagree (disacquiescence RS) with the items. In a 26-country study, participants who answered a survey in their first language tended to have a more extreme RS than participants who answered the survey in English (Harzing, 2006). However, English–Kannada bilinguals had a more extreme RS in English than in Kannada, although the association between English proficiency and extreme RS was weak (Messner, 2017). There are also cultural differences in acquiescence RS, people from individualistic cultures tend to show a less acquiescence RS (Johnson et al., 2005). However, it is unclear whether these cultural differences would generalize to languages. In short, the language of an instrument can influence answering tendencies and scores.

This study aims to examine whether the language of an instrument would influence the answers of Malaysian bilinguals or multilinguals. Due to its diverse racial and cultural background, the majority of the Malaysian population is bilingual or multilingual (Albury, 2017). In addition to the prominent use of Bahasa Malaysia and English, languages such as Mandarin and Tamil are used by the Chinese and Indian communities, respectively. This also echoes the country's education system; Bahasa Malaysia and English are taught and assessed formally in all primary and secondary schools. However, in the vernacular schools for the Chinese and Indian communities, in addition to Bahasa Malaysia and English, education is largely provided in Mandarin or Tamil, respectively (Y. S. Tan & Sezali, 2015). As a result, many Malaysians speak more than one language fluently. This provided the unique opportunity to measure possible language effects on instruments with a within-subject design.

Past studies in which language effects were observed (cultural accommodation or ethnic affirmation) typically used instruments measuring cultural values or variables that are known to be influenced by culture such as personality and attitudes. The cultural content of the questionnaires could hence explain any differences in scores or answering tendency. However, with this study, we aimed to find out whether language would also influence questionnaires with no specific cultural content, where one would hence not expect to find cultural differences. We used five questionnaires that do not explicitly examine cultural values or variables: The Adult Executive Functioning Inventory (ADEXI; Holst & Thorell, 2018), the Depression, Anxiety, and Stress Scale—21 items (DASS-21; Lovibond & Lovibond, 1995), the Brief Coping Orientation to Problems Experienced Inventory (Brief COPE; Carver, 1997), the Pediatric Quality of Life Inventory 4.0 (PedsQL; Varni et al., 2002), and the Traditional Masculinity-Femininity Scale (TMF; Kachel et al., 2016). If differences in these questionnaires were found, this could not be explained by cultural accommodation or ethnic affirmation, though a language might “trigger” a certain answering tendency.

The choice of these questionnaires was based on reflecting a broad scope of measures used in contemporary psychological studies: cognitive functioning (ADEXI), psychological distress (DASS), coping (Brief COPE), overall quality of life (PedsQL), and masculinity/femininity (TMF). We used broadly used and reliable measures, that were used in different countries, but language effects were not yet studied (see the “Methods” section for a broader description of these measures). We aimed to reflect a broad variety of psychological questionnaires, though needless to say, these choices were relatively arbitrary. Based on this study, we hope to get more insight into language influence on a broad scope of psychological topics that are seemingly culture-free.

We tested the difference between the answers on the questionnaires in Mandarin and English, or in Bahasa Malaysia and English. On our “culture-free” measures, we could, however, not predict the direction of a possible difference. If no differences were found between the languages, previous studies might be correct; the cultural content of the instruments might explain the language differences. However, if differences between languages were found, there could be two possible explanations: (1) The measures are not really “culture-free,” and hence the culture linked to a language

would influence the answers, even when not specifically measuring cultural values; or (2) the previously reported effects are not a true result of the cultural content of the questionnaires (language triggering cultural values which in turn influences responses) but might result from specific language factors, such as a different interpretation of the questions in different languages.

Differences between languages are one of the most likely contributors to how and why questions might be interpreted differently in different languages. Mandarin is a tonal language, and its orthography differs substantially from English, whereas Bahasa Malaysia overlaps with English to some degree. The overlap between Bahasa Malaysia and English is especially evident from the usage of the same Latin alphabet and the number of loanwords from English, such as *aktiviti* (activity), *kelas* (class), *mesej* (message), and so on. In contrast, Mandarin uses Chinese characters, which are logosyllabic, and each character represents one syllable of spoken Mandarin. Consequently, brain regions relevant to visual-orthographic processing are better developed in Mandarin speakers, and brain regions relevant to phonological processing are better developed in English speakers (Cao et al., 2015). However, Mandarin–English bilinguals, regardless of their proficiency, display mental representations of time that are comparable with Mandarin monolinguals (W. Yang et al., 2022). This suggests that one’s first language influences one’s cognition, in line with the linguistic relativity hypothesis, and further implies that cognition might not be reshaped by the acquisition of a second language. These differences or similarities in the underlying system of each language may, in turn, affect how people respond in each language.

Nonetheless, we expected to find a more Extreme RS in Mandarin and Bahasa Malaysia than in English since we explicitly recruited participants whose first language was Mandarin or Bahasa Malaysia, in line with Harzing (2006). Moreover, given the cultural differences in acquiescence RS (Harzing, 2006; Johnson et al., 2005), we explored whether the tendency to agree or disagree would also differ across languages. Given the exploratory nature of these analyses, no specific prediction was made.

Methods

Participants

To take part in the study, participants had to be fluent, that is, write and read, in English and Mandarin or in English and Bahasa Malaysia. Participants had to rate their proficiency and frequency of usage of the languages (English, Mandarin/Bahasa Malaysia) on an adapted Language History Questionnaire (LHQ; Li et al., 2014). The language use of Malaysians can be quite diverse (e.g., speaking different languages depending on the context, without having a very clear “first” or “second” language). To attain a relatively coherent sample, we explicitly aimed to collect data from participants who spoke either Mandarin or Bahasa Malaysia as their *first* language, and who also spoke English fluently. In the adapted LHQ, “first language” was inquired and defined as their mother tongue (native language), which is the language that they learned to speak first. We only included participants who indicated Mandarin or Bahasa Malaysia, respectively, for the Mandarin–English and Bahasa Malaysia–English samples.

Sample 1 Mandarin–English. A total of 85 Malaysian participants were recruited for the Mandarin/English sample. We aimed to compare the responses in the first (native) language Mandarin, and in English, which was explicitly mentioned in the recruitment material. Participants who did not indicate that Mandarin was their native language ($n = 10$) and who did not complete both questionnaires ($n = 10$) were excluded from the analyses. In addition, one participant was excluded due to

Table 1. Demographic variables (age/sex/ethnicity).

	Sample	
	Mandarin–English (N = 64)	Bahasa Malaysia–English (N = 47)
Sex		
Males	13 (20.3%)	7 (14.9%)
Females	50 (78.1%)	37 (78.7%)
Prefer not to say	1 (1.6%)	3 (6.4%)
Age		
Range	18–37	18–37
M ± SD	21.3 ± 2.9	22.4 ± 4.0
Ethnicity		
Malay	—	46 (97.9%)
Chinese	64 (100%)	—
Indian	—	—
Others	—	1 (2.1%)
Language proficiency M (SD)		
English	N = 64	N = 47
1. Reading	5.78 (0.83)	6.28 (0.85)
2. Writing	5.33 (0.94)	5.64 (1.03)
3. Speaking	5.25 (1.04)	5.74 (0.90)
4. Listening	5.67 (0.84)	5.87 (0.92)
Mandarin/Malay	N = 60	N = 39
1. Reading	6.37 (1.07)	6.59 (0.97)
2. Writing	5.77 (1.23)	5.78 (1.40)
3. Speaking	6.50 (0.85)	6.26 (1.07)
4. Listening	6.63 (0.74)	6.64 (0.74)
Mix language (% yes)	57 (89.1%)	46 (97.9%)
Which languages	37 (57.8%) Mandarin	12 (25.5%) Malay
Language think % (English/Mandarin/Malay)	16 (25.0%) English	26 (55.3%) English
	9 (14.1%) Both	9 (19.2%) Both
	2 (3.1%) Others	

age restriction (<18). This resulted in 64 participants (see Table 1 for demographic information) being included in the analyses.

Sample 2 Bahasa Malaysia–English. A total of 80 Malaysian participants were recruited for the Bahasa Malaysia/English sample. Given the objective to compare the responses in the first (native) language and English, participants who did not indicate that Bahasa Malaysia was their native language ($n=9$) and who did not complete both questionnaires ($n=24$) were excluded. This resulted in 47 participants (see Table 1 for demographic information) being included in the analyses.

The study was conducted in compliance with the University of Nottingham Code of Research Conduct and Research Ethics and approved by the University of Nottingham's Science and Engineering Research Ethics Committee (approval number: CZJ270720). Note that for an overarching project,

the relationships between measures were also evaluated, which are not included in the current manuscript.

Measures

Adult Executive Functioning Inventory. The ADEXI consists of 14 items which measure working memory (Items 1, 2, 5, 7, 8, 9, 11, 12, and 13) and inhibition (Items 3, 4, 6, 10, and 14). The ADEXI questionnaire is freely available in several languages (English, Swedish, Spanish, Spanish-Latino, Portuguese, Catalan, Traditional and Simplified Chinese, Polish, and Hungarian; see www.chexi.se). We used the available Chinese and Malay versions. Participants will rate the items on a 5-point rating scale from 1 = “*Definitely not true*” to 5 = “*Definitely true*.” The test–retest reliability and internal consistency of ADEXI were shown to be adequate, with bivariate correlations of .68–.72 (Holst & Thorell, 2018).

Brief Coping Orientation to Problems Experienced Inventory. The Brief COPE (Carver, 1997) consists of 28 items which measure 14 theoretically identified coping responses. The coping responses include so-called *approach* coping strategies, which include active coping (Items 2 and 7), emotional support (Items 5 and 15), use of informational support (Items 10 and 23), positive reframing (Items 12 and 17), planning (Items 14 and 25), and acceptance (Items 20 and 24), or conversely, *avoidant* coping strategies, which include self-distraction (Items 1 and 19), denial (Items 3 and 8), substance use (Items 4 and 11), behavioral disengagement (Items 6 and 16), venting (Items 9 and 21), and self-blame (Items 13 and 26). In addition, two coping responses that indicate neither approach nor avoidance coping strategies are humor (Items 18 and 28) and religion (Items 22 and 27). The Brief COPE has been translated into multiple languages such as Spanish, Greek, German, French, Korean, Chinese, and Malay. Participants will rate on a 4-point rating scale from 1 = “*I haven’t been doing this at all*” to 4 = “*I’ve been doing this a lot*.”

Depression, Anxiety, and Stress Scale—21 items. The DASS-21 (Lovibond & Lovibond, 1995) consists of 21 items which measure 3 broad emotion states: depression (Items 3, 5, 10, 13, 16, 17, and 21), anxiety (Items 2, 4, 7, 9, 15, 19, and 20), and stress (Items 1, 6, 8, 11, 12, 14, and 18). The DASS-21 questionnaire has been translated into multiple languages such as Chinese, Malay, and Tamil (<http://www2.psy.unsw.edu.au/dass/translations.htm>). Items are rated on a 4-point rating scale from 0 = “*Did not apply to me at all*” to 3 = “*Applied to me very much or most of the time*.”

Pediatric Quality of Life Inventory 4.0 (PedsQL 4.0 Generic Core Scale—adult version). The PedsQL (Varni et al., 2002) consists of 23 items that measure 4 categories: physical functioning, emotional functioning, social functioning, and school/work functioning. The PedsQL is available in Malaysian languages such as Malaysian English, Malaysian Mandarin Chinese, and Malaysian Tamil. Items are rated on a 5-point rating scale from 0 = “*never*” to 4 = “*almost always*.”

Traditional Masculinity-Femininity Scale. The TMF (Kachel et al., 2016) is a self-report questionnaire that consists of six items. The items measure the central facets of self-ascribed masculinity-femininity. The participants will be asked to rate themselves (e.g., attitude, behavior, and appearance) on a 7-point rating scale from 1 = “*very masculine*” to 7 = “*very feminine*.” Although *what* is considered masculine or feminine might be influenced by culture, the current scale merely measures *how* masculine/feminine participants rate themselves and is hence relatively culture-free. The TMF was not available in Mandarin and Bahasa Malaysia and hence was translated in line with translation guidelines (Tsang et al., 2017), with a forward and back-translation procedure.

Procedure

Participants provided informed consent before participating in the study. After providing demographic information, and filling in the language proficiency questionnaire, they were asked to fill in the five questionnaires (ADEXI, DASS-21, Brief COPE, PedsQL, and TMF). The order of the language of the questionnaires was counterbalanced within both samples. Half of the samples completed the English version first, while the other half completed the Mandarin/Bahasa Malaysia version first. The sequence of the questionnaires was also randomized across participants. The follow-up questionnaire in English, Mandarin, or Bahasa Malaysia was sent to participants approximately 1 week after they completed the first questionnaire. If the participants did not respond to the follow-up questionnaire, up to three reminder emails were sent. The study took approximately 20 minutes for each session. The participants were given course credits or compensation by a lucky draw prize upon completion of the study.

Results

Analyses

To test whether there was an overall language difference on questionnaires, total scores were calculated for the ADEXI, TMF, and PedsQL. On the Brief COPE, a total score cannot be calculated, hence each coping strategy was calculated (sum of two items) separately (see Table 3). For the DASS, the three subscales were calculated (see Table 3). Extreme RS was calculated by counting the number of extreme responses on each questionnaire (see Table 2). The acquiescence RS was calculated as a percentage of items that a participant agrees with.¹

To compare questionnaire scores, extreme, and acquiescence RS on each questionnaire, paired-sample *t*-tests or Wilcoxon signed-rank tests were conducted depending on the type of data (i.e., continuous or count) and whether the data were normally distributed. The significance level was set to be at $p < .0017$ to account for multiple comparisons (Bonferroni correction).

Mandarin versus English

Scores were significantly higher on the Mandarin version than on the English version of several subscales of the Brief COPE, the Stress subscale of the DASS, and the total PedsQL (see Table 3 for statistical data). On the other measures, there were no significant differences between Mandarin and English.

Participants provided more extreme responses in Mandarin than in English, and a more acquiescence RS in English than in Mandarin on the PedsQL. Note that the PedsQL is reversely transformed and scored; agreeing more with a statement would lead to a *lower* score. On the Brief COPE, participants showed more acquiescence RS in Mandarin than in English. On the other questionnaires, there were no significant differences in extreme or acquiescence RS (see Table 4).

Malay versus English

None of the questionnaires differed between the Bahasa Malaysia and English $ps > .001$ (see Table 3). However, participants provided more extreme responses in English than Bahasa Malaysia on the Brief COPE (see Table 4). In addition, participants showed a more acquiescence RS in Bahasa Malaysian than in English. On the other questionnaires, there were no significant differences in Extreme, Middle, or Acquiescence RS (see Table 4).

Table 2. The rating scale structure of each questionnaire.

Measure	Scale	Extreme (E)	Acquiescence (A) versus Disacquiescence (D)
ADEXI	1 <i>Definitely not true</i>	E	D
	2 <i>Not true</i>		D
	3 <i>Partially true</i>		A
	4 <i>True</i>		A
	5 <i>Definitely true</i>	E	A
Brief COPE	1 <i>I haven't been doing this at all</i>	E	D
	2 <i>A little bit</i>		A
	3 <i>A medium amount</i>		A
	4 <i>I've been doing this a lot</i>	E	A
DASS-21	0 <i>Did not apply to me at all</i>	E	D
	1 <i>Applied to me to some degree or some of the time</i>		A
	2 <i>Applied to me to a considerable degree or a good part of the time</i>		A
	3 <i>Applied to me very much or most of the time</i>	E	A
TMF	1 <i>Very masculine</i>	E	
	7 <i>Very feminine</i>	E	
PedsQL	0 <i>Never</i>	E	D
	1 <i>Almost never</i>		D
	2 <i>Sometimes</i>		A
	3 <i>Often</i>		A
	4 <i>Almost Always</i>	E	A

Note. E, extreme; M, middle; A, acquiescence; D, disacquiescence.

Discussion

The language of a questionnaire is known to influence the RS on culture-related questionnaires. With this study, we aimed to explore whether language would also influence answers on questionnaires with no specific cultural content. In this within-subject study, we looked at the differences between Mandarin or Bahasa Malaysia versus English in scores and RSs on five psychological questionnaires (ADEXI, Brief COPE, DASS, TMF, and PedsQL). Although there were no differences on the ADEXI and TMF in scores and RS, the other questionnaires did show some differences. The differences between languages indicate that either these measures are not culture-free, and hence cultural accommodation might take place, or previously found language differences did not fully result from culture, and other factors such as translation or interpretation might explain the difference.

Mandarin versus English

Coping. On the Brief COPE subscales Emotional Support, Informational Support, Planning, Acceptance, and Self-Blame scores were higher in Mandarin than in English. Participants also had a more acquiescence RS in Mandarin, which might (partly) explain the higher scores in Mandarin.

However, one could argue that coping mechanisms are influenced by culture. For instance, the Sociocultural Stress and Coping Model highlighted the role of culture in determining coping styles

Table 3. Comparisons of the means (SD) for t-tests /medians (range) for Wilcoxon tests between Mandarin/Bahasa Malaysia and English.

	Mandarin	English	Test & value	p	BM	English	Test and value	p
ADEXI	40.11 (8.48)	38.92 (8.28)	t(63) = -1.73	.09	40.53 (10.4)	41.04 (9.5)	t(46) = 0.54	.59
COPE								
Active Coping	6 (3-8)	6 (2-8)	Z = -2.17	.03	6 (4-8)	6 (2-8)	Z = -0.03	.97
Emotional Support	5 (2-8)	4 (2-8)	Z = -3.52	<.001	5 (2-8)	6 (2-8)	Z = -0.22	.83
Informational Support	6 (2-8)	5 (2-8)	Z = -4.10	<.001	6 (2-8)	5 (2-8)	Z = -2.17	.03
Positive Reframing	6 (3-8)	5 (2-8)	Z = -1.73	.09	6 (2-8)	6 (3-8)	Z = -0.96	.34
Planning	6 (4-8)	5.5 (2-8)	Z = -3.54	<.001	6 (4-8)	6 (2-8)	Z = -0.59	.56
Acceptance	7 (3-8)	6 (2-8)	Z = -3.33	<.001	6 (3-8)	6 (2-8)	Z = -1.51	.13
Self-Distraction	6 (2-8)	5 (3-8)	Z = -1.49	.14	6 (3-8)	6 (2-8)	Z = -1.08	.28
Denial	3 (2-6)	3 (2-5)	Z = -2.50	.01	4 (2-7)	4 (2-7)	Z = -0.20	.84
Substance Use	2 (2-8)	2 (2-8)	Z = -1.27	.21	2 (2-6)	2 (2-8)	Z = -1.27	.21
Behavioral Disengagement	4 (2-7)	4 (2-8)	Z = -0.73	.46	4 (2-7)	4 (2-7)	Z = -0.66	.51
Venting	5 (2-8)	4 (2-8)	Z = -2.75	.01	5 (2-8)	5 (2-8)	Z = -0.68	.49
Self-Blame	6 (3-8)	5 (2-8)	Z = -3.74	<.001	6 (2-8)	6 (2-8)	Z = -1.72	.09
Humor	4 (2-8)	4 (2-8)	Z = -1.93	.05	4 (2-8)	4 (2-8)	Z = -0.53	.60
Religion	4 (2-8)	3.5 (2-8)	Z = -2.58	.01	6 (2-8)	6 (2-8)	Z = -0.74	.46
DASS								
Depression	14 (0-38)	12 (2-38)	Z = -0.99	.32	16 (0-42)	12 (2-42)	Z = -0.01	1.00
Anxiety	10 (0-38)	12 (0-40)	Z = -0.31	.76	12 (0-42)	12 (0-38)	Z = -0.30	.76
Stress	18 (0-42)	14 (0-36)	Z = -3.98	<.001	18 (0-40)	14 (0-40)	Z = -0.57	.57
TMF	26.31 (6.25)	26.36 (5.66)	t(63) = 0.12	.91	27.15 (6.99)	27.47 (7.24)	t(46) = 0.64	.53
PedsQL	1,543.75 (336.44)	1,455.08 (295.31)	t(63) = -4.28	<.001	1,398.40 (353.11)	1,343.62 (350.75)	t(46) = -1.80	.08

Note. BM, Bahasa Malaysia.

Table 4. Comparisons of median (range) RS on each questionnaire across languages.

Sample	English		p	BM	English		p
	Mandarin	English			Test and value	English	
ADEXI							
Extreme RS	1.5 (0-8)	1 (0-7)	.84	3 (0-13)	2 (0-10)	Z=-0.38	.70
Acquiescence RS	0.64 (0.07-1.00)	0.57 (0.14-1.00)	.25	0.64 (0.00-1.00)	0.71 (0.07-1.00)	Z=-1.15	.49
COPE							
Extreme RS	11.5 (1-18)	10.5 (2-21)	.72	9 (1-25)	13 (2-23)	Z=-0.36	<.001
Acquiescence RS	0.82 (0.57-1.00)	0.75 (0.25-0.96)	<.001	0.86 (0.57-1.00)	0.82 (0.18-0.96)	Z=-3.43	<.001
DASS							
Extreme RS	10 (2-20)	8 (0-19)	.21	9 (1-21)	10 (1-18)	Z=-1.27	.82
Acquiescence RS	0.71 (0.05-1.00)	0.71 (0.10-1.00)	.52	0.67 (0.00-1.00)	0.71 (0.14-1.00)	Z=-0.64	.97
TMF							
Extreme RS	0 (0-6)	0 (0-3)	.12	0 (0-5)	0 (0-6)	Z=-1.57	.57
PedsQL							
Extreme RS	7.5 (0-23)	5 (0-21)	<.001	7 (1-19)	7 (0-18)	Z=-3.58	.90
Acquiescence RS	0.41 (0.00-0.83)	0.48 (0.04-0.87)	<.001	0.57 (0.09-0.87)	0.52 (0.17-0.91)	Z=-4.67	.24

Note. The TMF does not imply a tendency to agree or disagree. BM, Bahasa Malaysia.

(Knight & Sayegh, 2010). Certain cultures might promote certain coping strategies. Asian American students, with a collectivistic cultural background, might be more pessimistic than Caucasian American students with an individualistic cultural background (Chang, 1996). This, in turn, might increase negative problem orientation (Chang, 1996). In addition, comparisons between Taiwanese, Chinese, and South Korean Asian international students from collectivistic cultures showed differences in coping strategies (Constantine et al., 2004). Therefore, cultural accommodation might lead to higher scores on those coping strategies that are promoted in the culture linked to the language of the questionnaire. Although this suggests that cultural accommodation took place, there was no higher score in coping strategies in Bahasa Malaysia compared with English.

DASS. Scores on the Stress subscale of the DASS were higher in Mandarin than in English, which could not be explained by RS. This could indicate that stress is culture-dependent. If stress is more common in the Chinese/Malaysian culture, higher scores might be explained by cultural accommodation. Academic expectations are the most common triggers of stress among university students. In the Asian Confucian Heritage Culture, academic expectations are high, and this is strongly related to stress among students (J. B. Tan & Yates, 2011). Students might link higher academic expectations related stress to the Mandarin language and hence score higher in Mandarin (i.e., cultural accommodation).

Stigma regarding mental health issues might have influenced responses on the DASS in all languages. In many Asian countries, there is a prevailing belief that mental health problems should be hidden or denied to protect the family's honor and social standing (Gopalkrishnan, 2018; Zhang et al., 2020). As a result, individuals experiencing mental health issues often face pressure to conform to societal norms and delay help-seeking (Javed et al., 2021). However, there is growing awareness and advocacy in Asia to challenge these stereotypes and promote mental health education and support, aiming to reduce the stigma associated with these issues and create more inclusive and compassionate societies (Zhang et al., 2020).

PedsQL. The PedsQL scores were higher (i.e., better Quality of Life) in Mandarin than in English. This difference could be explained by the RS; participants gave more extreme responses in Mandarin which might lead to a higher score and more agreeable responses in English which might lead to a lower score. From previous studies, it is known that people tend to give more extreme answers in the language they are most fluent in. In our Chinese sample, Mandarin was their first language, hence one might expect that participants were more fluent (which, when eyeing the data, seemed true). Most participants indicated that they think in Mandarin. This could hence explain why more extreme answers were given in Mandarin. Cross-cultural or -language differences in the quality of life measure might also explain the discrepancy in scores. English adolescents with and without Cystic Fibrosis reported poorer quality of life than their German peers despite their matched functioning, suggesting an influence of culture or language on reports of quality of life (Abbott et al., 2001). It seems that a combination of RS and cross-cultural/-language differences in quality of life measures may explain the difference.

Alternatively, the lower score in English than Mandarin PedsQL may reflect a less biased self-report of quality of life. Participants showed reduced decision biases when using a foreign language compared with their first language (Keysar et al., 2012). By the same token, our participants might have provided more unbiased judgments of their own quality of life in English, possibly due to the greater cognitive and emotional distance offered by a foreign language. We are unable to confirm if this is the case given that quality of life is fundamentally a subjective measure, but future studies could investigate whether using a first and second language deviates from an objective measure.

Bahasa Malaysia versus English

None of the questionnaire scores differed between Bahasa Malaysia and English. However, on the Brief COPE, participants gave more extreme responses in English than in Bahasa Malaysia, while they gave more agreeable responses in Bahasa Malaysia than in English. Even though we explicitly invited participants who had Bahasa Malaysia as their primary language, in our Bahasa Malaysia sample, the majority of the participants indicated that they normally *think* in English (see Table 1). Possibly, English is hence more prominent in their mind and might lead to more extreme answers. Moreover, from a cultural perspective, one could argue that in Western cultures (linked to the English language) a more extreme and less agreeable RS could be expected (Harzing, 2006; Johnson et al., 2005), which aligns with these findings.

Above we have given several possible explanations about how culture might still be involved in answering these seemingly culture-free questionnaires. However, alternatively, the interpretation of the questions might be different in Mandarin or Bahasa Malaysia than in English. Several terms might have a different connotation in a different language. For example, while the English word “spontaneous” in the autism spectrum quotient broadly refers to doing something without prior planning, the Bengali and Hindi words for “spontaneous” align closer with “taking initiative” (Carruthers et al., 2018). Moreover, the evaluation (whether the word/behavior has a positive or negative value) of such words might be influenced by this connotation in a certain culture, such as differences in uncertainty avoidance, power distance, individualism, or collectivism (Hofstede, 2001). In short, culture might also influence a broader interpretation of words, even when those words do not intend to measure a specific cultural construct. Since all our measures included psychological constructs and behaviors, one could imagine that the terms for these constructs and behaviors are evaluated differently in different cultures (e.g., some terms might be more stigmatized). Moreover, there are indications that the standard back-translation may not guarantee equivalence in interpretations across languages (Barger et al., 2010). This shows that besides back-translation, cultural validation is needed. Moreover, our results show that even *within* the same culture, the language can already influence questionnaire scores.

TMF and ADEXI

Both in the Mandarin and Bahasa Malaysia samples, there were no differences between languages in scoring nor RS on the TMF and ADEXI. This might not be surprising, given that such differences in previous studies were mostly found in questionnaires with cultural content. The questions in the ADEXI cover Executive Functions, which are culturally neutral questions about inhibition and working memory. We would not expect any cultural or language influences on these measures. It is slightly more surprising that we did not find any differences in the TMF. Although the content of the TMF is not explicitly cultural, the interpretation of masculinity and femininity might be interpreted and evaluated differently in different cultures. What is considered to be feminine or masculine traits in one culture might not be universal across cultures (Ward & Sethi, 1986). However, the currently used questionnaire does not measure specific masculine and feminine *traits* but merely asks whether one considers themselves to be masculine/feminine. Hence such specific differences in interpretation would not be reflected on this questionnaire. This might hence explain why there was no difference in any of the measures (scores, acquiescence or extreme RS).

General discussion

Although there were some differences between languages, these differences were relatively small. An explanation could be that in the current sample, the different languages did not strongly trigger

different cultural values. Previous studies on language often included minority populations, who spoke the main language of the country they lived in (i.e., English in the United States) and their mother tongue (e.g., Chinese) at home. In such conditions, the “home” language is linked to an intrinsically different culture than the country’s language and culture (e.g., being Chinese in the United States). However, in the current sample, the cultural values of the different languages might not be very different. Although Malaysia is the home to a variety of ethnicities, each with their unique cultural values, these ethnicities have been living in Malaysia for many generations. Hence, apart from the unique ethnic culture, they share a common Malaysian culture. Both English, Bahasa Malaysia and Chinese are inherent languages of the Malaysian culture, and hence the difference between the culture linked to one’s “home” language, or mother tongue, might not be that different from the country’s culture and language. This is also reflected in the number of participants mixing languages, which is very common in Malaysia (also called *Bahasa Rojak* “language salad”). In short, the different languages that people speak in Malaysia might not be consistently attached to an inherently different culture.

Language is an important determinant of RS (Harzing, 2006), but differences in RS between languages were only found on the Brief COPE and PedsQL. The lack of differences in RS between languages on other measures may be due to the variance in the wording or nature of the Likert-type scales. Harzing and colleagues (2009) reported that a 7-point scale diminished extreme RS to a certain extent in comparison to a 5-point scale. This suggests that participants might be able to qualify their responses better with a broader scale, and the lack of differences on TMF specifically appears to support this. The perceived severity of the response labels on a quality of life measure differed between Mandarin, and Bahasa Malaysia, indicating that concept or semantic equivalence may not necessarily guarantee measurement or interpretation invariance (Luo et al., 2015). Therefore, it is likely that the interpretation of Likert-type scales differs between languages, especially on the Brief COPE and PedsQL, whereas the Likert-type scale of other measures might be interpreted similarly across languages.

Caveats

This study looked at the influence of language on questionnaires with no specific cultural content. Although we used a within-subject design, which gives important insight into the possible influence that language has within individuals, and we used well-translated questionnaires, there were some caveats. First, there were more females than males in the samples and a relatively young age range. Although we do not expect that language would have a different influence on questionnaires in males and females, or on a specific age range, this does mean that the current findings cannot be generalized to the broader general population. Second, apart from differences in scores and RS, differential item functioning, measurement invariance, and item response theory could give insight into a questionnaire functioning. In this study, our numbers did not allow us to do such analyses, but this is highly recommended for future research. Third, the sample size was unequal between Mandarin–English and Bahasa Malaysia–English speakers, and both samples were relatively small. Moreover, there is the possibility that English is a third language rather than the second language of our participants. However, we believe this is highly unlikely given the educational landscape of Malaysia where English is usually taught from a very young age in both national and vernacular schools. Finally, although out of the scope of this study, to measure possible cultural bias in the questionnaires, future studies could (1) gather cultural background information about participants such as ethnicity, or birth country, or (2) add measures of culture (e.g., Hofstede, 1984), such as collectivism, or power distance, to check whether these factors are related to the scores on the currently used questionnaires.

Conclusion

Our findings indicate that the ADEXI and TMF can be used in Bahasa Malaysia, English, and Mandarin, as there are no differences in scores or RS. However, scores on the Brief COPE, DASS stress scale, and PedsQL might be influenced by language. This indicates that even on questionnaires with no specific cultural content, cultural accommodation might take place and RS might be influenced by language (proficiency). This indicates that apart from proper translation, questionnaires should be validated in different languages, even if administered in the same country. This is relevant for clinical practice in highly multilingual countries; often questionnaires are provided in different translations, but the language might influence the scores.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Yee Thung Lee  <https://orcid.org/0000-0002-0363-7352>

Marieke de Vries  <https://orcid.org/0000-0003-2845-8956>

Note

1. Disacquiescence and Middle RS were not analyzed, as these are the inverse of the Acquiescence and Extreme RS, respectively (see Table 2).

References

- Abbott, J., Baumann, U., Conway, S., Etherington, C., Gee, L., Von Der Schulenburg, J. M. G., & Webb, K. (2001). Cross cultural differences in health related quality of life in adolescents with cystic fibrosis. *Disability and Rehabilitation*, 23(18), 837–844. <https://doi.org/10.1080/09638280110072913>
- Albury, N. J. (2017). Mother tongues and languaging in Malaysia: Critical linguistics under critical examination. *Language in Society*, 46(4), 567–589. <https://doi.org/10.1017/S0047404517000239>
- Barger, B., Nabi, R., & Hong, L. Y. (2010). Standard back-translation procedures may not capture proper emotion concepts: A Case study of Chinese disgust terms. *Emotion*, 10(5), 703–711. <https://doi.org/10.1037/a0021453>
- Baumgartner, H., & Steenkamp, J. B. E. M. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research*, 38(2), 143–156. <https://doi.org/10.1509/jmkr.38.2.143.18840>
- Bond, M. H., & Yang, K. S. (1982). Ethnic affirmation versus cross-cultural accommodation: The variable impact of questionnaire language on Chinese bilinguals from Hong Kong. *Journal of Cross-Cultural Psychology*, 13(2), 169–185. <https://doi.org/10.1177/0022002182013002003>
- Cao, F., Brennan, C., & Booth, J. R. (2015). The brain adapts to orthography with experience: Evidence from English and Chinese. *Developmental Science*, 18(5), 785–798.
- Carruthers, S., Kinnaird, E., Rudra, A., Smith, P., Allison, C., Auyeung, B., Chakrabarti, B., Wakabayashi, A., Baron-Cohen, S., Bakolis, I., & Hoekstra, R. A. (2018). A cross-cultural study of autistic traits across India, Japan and the UK. *Molecular Autism*, 9(1), 1–10. <https://doi.org/10.1186/s13229-018-0235-3>
- Carver, C. S. (1997). You want to measure coping but your protocol's too long: Consider the brief COPE. *International Journal of Behavioral Medicine*, 4(1), 92–100.

- Chang, E. (1996). Evidence for the cultural specificity of pessimism in Asians vs Caucasians: A test of a general negativity hypothesis. *Personality and Individual Differences, 21*(5), 819–822. [https://doi.org/10.1016/0191-8869\(96\)00110-9](https://doi.org/10.1016/0191-8869(96)00110-9)
- Chee, Z. J., & de Vries, M. (2021). Language matters : The Autism-Spectrum Quotient in English, Mandarin and Bahasa Malaysia. *Journal of Autism and Developmental Disorders, 52*, 3814–3824. <https://doi.org/10.1007/s10803-021-05253-9>
- Chen, S. X., Benet-Martínez, V., & Ng, J. C. K. (2014). Does language affect personality perception? A functional approach to testing the Whorfian hypothesis. *Journal of Personality, 82*(2), 130–143. <https://doi.org/10.1111/jopy.12040>
- Chen, S. X., & Bond, M. H. (2007). Explaining language priming effects: Further evidence for ethnic affirmation among Chinese-English bilinguals. *Journal of Language and Social Psychology, 26*(4), 398–406.
- Constantine, M. G., Okazaki, S., & Utsey, S. O. (2004). Self-concealment, social self-efficacy, acculturative stress, and depression in African, Asian, and Latin American International College Students. *American Journal of Orthopsychiatry, 74*(3), 230–241.
- Gopalkrishnan, N. (2018). Cultural diversity and mental health: Considerations for policy and practice. *Frontiers in Public Health, 6*, Article 179. <https://doi.org/10.3389/fpubh.2018.00179>
- Harzing, A. W. (2006). Response styles in cross-national survey research: A 26-country study. *International Journal of Cross Cultural Management, 6*(2), 243–266. <https://doi.org/10.1177/1470595806066332>
- Harzing, A. W., Baldueza, J., Barner-Rasmussen, W., Barzantny, C., Canabal, A., Davila, A., Espejo, A., Ferreira, R., Giroud, A., Koester, K., Liang, Y. K., Mockaitis, A., Morley, M. J., Myloni, B., Odusanya, J. O. T., O’Sullivan, S. L., Palaniappan, A. K., Prochno, P., Choudhury, S. R., . . . Zander, L. (2009). Rating versus ranking: What is the best way to reduce response and language bias in cross-national research? *International Business Review, 18*(4), 417–432. <https://doi.org/10.1016/j.ibusrev.2009.03.001>
- Hofstede, G. (1984). *Culture’s consequences: International differences in work-related values* (Vol. 5). Sage.
- Hofstede, G. (2001). *Culture’s consequences: Comparing values, behaviors, institutions and organizations across nations*. Sage.
- Holst, Y., & Thorell, L. B. (2018). Adult executive functioning inventory (ADEXI): Validity, reliability, and relations to ADHD. *International Journal of Methods in Psychiatric Research, 27*(1), Article e1567. <https://doi.org/10.1002/mpr.1567>
- Javed, A., Lee, C., Zakaria, H., Buenaventura, R. D., Cetkovich-Bakmas, M., Duailibi, K., Ng, B., Ramy, H., Saha, G., Arifeen, S., Elorza, P. M., Ratnasingham, P., & Azeem, M. W. (2021). Reducing the stigma of mental health disorders with a focus on low-and middle-income countries. *Asian Journal of Psychiatry, 58*, 102601. <https://doi.org/10.1016/j.ajp.2021.102601>
- Johnson, T., Kulesa, P., Cho, Y. I., & Shavitt, S. (2005). The relation between culture and response styles: Evidence from 19 countries. *Journal of Cross-Cultural Psychology, 36*(2), 264–277. <https://doi.org/10.1177/0022022104272905>
- Kachel, S., Steffens, M. C., & Niedlich, C. (2016). Traditional masculinity and femininity: Validation of a new scale assessing gender roles. *Frontiers in Psychology, 7*, 1–19. <https://doi.org/10.3389/fpsyg.2016.00956>
- Kemmelmeier, M., & Cheng, B. Y. M. (2004). Language and self-construal priming: A replication and extension in a Hong Kong sample. *Journal of Cross-Cultural Psychology, 35*(6), 705–712. <https://doi.org/10.1177/0022022104270112>
- Keysar, B., Hayakawa, S. L., & An, S. G. (2012). The foreign-language effect: Thinking in a foreign tongue reduces decision biases. *Psychological Science, 23*(6), 661–668.
- Knight, B. G., & Sayegh, P. (2010). Cultural values and caregiving: The updated sociocultural stress and coping model. *Journals of Gerontology—Series B Psychological Sciences and Social Sciences, 65* B(1), 5–13. <https://doi.org/10.1093/geronb/gbp096>
- Li, P., Zhang, F., Tsai, E., & Puls, B. (2014). Language History Questionnaire (LHQ 2.0): A new dynamic web-based research tool. *Bilingualism, 17*(3), 673–680. <https://doi.org/10.1017/S1366728913000606>
- Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour Research and Therapy, 33*(3), 335–343. [https://doi.org/10.1016/0005-7967\(94\)00075-U](https://doi.org/10.1016/0005-7967(94)00075-U)

- Luo, N., Wang, Y., How, C. H., Tay, E. G., Thumboo, J., & Herdman, M. (2015). Interpretation and use of the 5-level EQ-5D response labels varied with survey language among Asians in Singapore. *Journal of Clinical Epidemiology*, *68*(10), 1195–1204. <https://doi.org/10.1016/j.jclinepi.2015.04.011>
- Messner, W. (2017). Market research in India: Does the choice of language cause questionnaire contamination? *Journal of Indian Business Research*, *9*(2), 149–168. <https://doi.org/10.1108/JIBR-09-2016-0100>
- Peytcheva, E. (2018). Can the language of survey administration influence respondents' answers? In T. P. Johnson, B.-E. Pennell, I. A. L. Stoop, & B. Dorer (Eds.), *Advances in Comparative Survey Methods* (pp. 325–340). Wiley. <https://doi.org/10.1002/9781118884997.ch15>
- Richard, M. O., & Toffoli, R. (2009). Language influence in responses to questionnaires by bilingual respondents: A test of the Whorfian hypothesis. *Journal of Business Research*, *62*(10), 987–994. <https://doi.org/10.1016/j.jbusres.2008.10.016>
- Tan, J. B., & Yates, S. (2011). Academic expectations as sources of stress in Asian students. *Social Psychology of Education*, *14*(3), 389–407. <https://doi.org/10.1007/s11218-010-9146-7>
- Tan, Y. S., & Sezali, M. D. (2015). *The emergence and widening of ethnic divide in the Malaysian educational system*. Strategic Information and Research Development Centre.
- Tsang, S., Royse, C. F., & Terkawi, A. S. (2017). Guidelines for developing, translating, and validating a questionnaire in perioperative and pain medicine. *Saudi Journal of Anesthesia*, *11*(Suppl. 1), S80. <https://doi.org/10.4103/sja.SJA>
- Varni, J. W., Seid, M., Knight, T. S., Uzark, K., & Szer, I. S. (2002). The PedsQLTM 4.0 Generic Core Scales: Sensitivity, Responsiveness, and impact on clinical decision-making. *Journal of Behavioral Medicine*, *25*(2), 175–193. <https://doi.org/10.1023/A:1014836921812>
- Ward, C., & Sethi, R. R. (1986). Cross-Cultural Validation of the Bem Sex Role Inventory: Malaysian and South Indian research. *Journal of Cross-Cultural Psychology*, *17*(3), 300–314.
- Yang, K. S., & Bond, M. H. (1980). Ethnic affirmation By Chinese Bilinguals. *Journal of Cross-Cultural Psychology*, *11*(4), 411–425. <http://hjb.sagepub.com.proxy.lib.umich.edu/content/9/2/183.full.pdf+html>
- Yang, W., Gu, Y., Fang, Y., & Sun, Y. (2022). Mental representations of time in English monolinguals, Mandarin monolinguals, and Mandarin–English bilinguals. *Frontiers in Psychology*, *13*, Article 791197.
- Zhang, Z., Sun, K., Jatchavala, C., Koh, J., Chia, Y., Bose, J., Li, Z., Tan, W., Wang, S., Chu, W., Wang, J., Tran, B., & Ho, R. (2020). Overview of stigma against psychiatric illnesses and advancements of anti-stigma activities in six Asian societies. *International Journal of Environmental Research and Public Health*, *17*(1), 280. <https://doi.org/10.3390/ijerph17010280>

Author biographies

Zhong Jian Chee is a Postdoctoral researcher at the School of Psychology, University of Aberdeen, studying the effects of unpleasant-to-touch materials/objects on actions and behaviors. He worked on the current project during his PhD at the School of Psychology, University of Nottingham Malaysia.

Yee Thung Lee is a PhD student at the Institute for Transport Studies, University of Leeds, studying attention during driving with a driving simulator and eye tracking. She worked on the current project as a Research Assistant at the School of Psychology, University of Nottingham Malaysia.

Omid R. Fani worked on the current project during his PhD at the School of Psychology, University of Nottingham Malaysia.

Tze Wei Yong worked on this project as a Research Assistant at the School of Psychology, University of Nottingham Malaysia.

Marieke de Vries is an Assistant Professor at the Department of Development and Education of Youth in Diverse Societies (DEEDS) at Utrecht University, and an Honorary Associate Professor at the School of Psychology, University of Nottingham Malaysia. Her research focuses on Autism, Culture, and Executive Functioning. The current project was part of a research series into questionnaire bias in cross-cultural research.