JOURNAL OF COMPUTER-MEDIATED COMMUNICATION

Short and Sweet? Length and Informative Content of Open-Ended Responses Using SMS as a Research Mode

Erin Walsh

Centre for Research on Ageing, Health and Wellbeing, Research School of Population Health, Australian National University, 54 Mills Road, Acton, ACT 2601

Jay K Brinker

Psychological Sciences, Swinburne University of Technology, Hawthorn, Vic 3122

Short Message Service (SMS) is one of the most widely used data services worldwide. This paper examines the assumption that the 160-character limit would force brief and thus comparatively uninformative responses in psychological research compared to other data collection modes. In laboratory classes, 463 psychology undergraduate students were randomly assigned to complete a 2-item questionnaire by SMS, e-mail, online survey, or paper survey. 2 weeks later, participants completed a multiple-choice self-report risk taking questionnaire on paper. While SMS response lengths were statistically significantly shorter than those yielded in other modes, they did not contain less information.

Keywords: Text Messaging, Mode, Response Properties, Methodology.

doi:10.1111/jcc4.12146

One ofthe most widely used data services worldwide (Kuntsche & Robert, 2009), 36.3 billion Short Message Service (SMS) were sent in 2011 in Australia (ACMA, 2011). As mobile telephones are ubiquitous (Anhoj & Moldrup, 2009), and their SMS capabilities are used daily by the majority of people (Mackay & Weidlich, 2009), they provide unprecedented opportunities for researchers to communicate with participants wherever they may be, at any time of day (Haller, Sanci, Sawyer, Coffey, & Patton, 2006). SMS is an attractive option as commercial bulk text services make bidirectional SMS communication with large groups increasingly affordable (Steeh, Buskirk, & Callegaro, 2007). It is also pragmatically feasible as many of these services offer automatic response aggregation just as online survey services do, which avoids the cost and potential for error incumbent in data entry from modes such as paper responses (Johansen & Wedderkopp, 2010; Tomlinson et al., 2009). SMS constitutes an exciting opportunity for real-time, bidirectional communication between researchers and participants, however there is limited rigorous, methodical examination ofits properties as a tool for psychological research (Cocco & Tuzzi, 2012; Tomlinson et al., 2009). Just as aspects of questionnaires other than their psychological content

Editorial Record: First manuscript received on June 4, 2014. Revisions received on October 13, 2014, March 23, 2015 and May 22, 2015. Accepted by Matthew Lombard on July 9, 2015. Final manuscript received on July 25, 2015. First published online on January 13, 2016.

can impact on participant responses (e.g. question wording), the mode used in psychological research can impact on how participants engage with and respond to a questionnaire (Cocco & Tuzzi, 2012; Dillman, Smyth, & Christian, 2009; Kelly, 2002; Lynn & Kaminska, 2012). One important issue is whether SMS gathers less useful information than other modes. The character-length restriction and cost of each SMS may limit the number of questions a researcher may ask, as well as length of responses. The two are interrelated because of the character limit of SMS, as more questions leave less room for responses to each question. In the interests of parsimony the current paper will focus on response length when the number of questions asked are fixed.

Open-ended questions do not have a predefined response set, such as yes / no or numbers in Likert-type scales, leaving the participant free to respond however they see fit. When constructing an open-ended question, the researcher should consider the length of response that they wish participants to give. This is related to pragmatic concerns, as shorter answers tend to be easier to categorise and can be more quickly analysed, while longer answers can provide more in-depth information by allowing participants more space to express their response. While there is some research regarding response length in the context of traditional data collection modes such as paper or online surveys (Kelly, Harper, & Landau, 2008), guidelines that apply to those modes do not necessarily translate well to SMS. In paper and online surveys, the amount of space afforded to the participant to answer the question can act as a subtle cue of researcher intent (Dillman et al., 2009). The amount of space allowed for responses is easily altered through adding carriage returns on paper surveys, and most online survey services offer at least rudimentary control over the size of text boxes intended for open-ended responses. However, no such control can be offered by SMS.

SMS sent in everyday life may be an indicator of what length may be expected from open-ended questions sent via SMS. Two often mentioned reasons for brevity in SMS are the character limit, and the social use of abbreviations (Doring, 2002). While it may be intuitively expected that the 160-character limit of SMS might foster brief responses, most mobile networks offer seamless stitching of multiple SMS into a single message (linked SMS) making this limit less salient to individuals sending SMS. Whether on an 'unlimited' or fixed credit plan, in everyday SMS usage, people do not keep track of the length of their SMS messages (Battestini, Setlur, & Sohn, 2010), suggesting they do not actively attempt to minimise conversations to fit in a single text. Frehner and Lang (2008) found every day, unlinked SMS communication was under the 160 character limit, with an average of approximately 20 words, or 97 characters. Linked SMS, however, were longer, containing an average of approximately 50 words, or 280 characters. The longest linked SMS in the corpus was 94 words, or 485 characters. It should be noted that 485 characters is not neatly divisible by 160, indeed only six characters less would have made the linked message the equivalence in cost to three individual messages, rather than four. This suggests that people may not be paying attention to the character limits of their messages to minimise costs. To place these message lengths in context, email messages within the same corpus were an average of 131 words, containing up to 1192 characters.

Cocco and Tuzzi (2012) proposed that SMS and face-to-face interview data are similar as SMS engenders a 'virtual presence' of the researcher, resulting in interview and SMS data involving a similar social, synchronous, bidirectional element to communication between the researcher and participant, which is lacking in other research modes such as online or paper surveys. Indeed, the main reason for sending an SMS is for social contact (Battestini et al., 2010; Polite, 2001). In a general social usage it is not the content of the message, but the gesture of sending the message that is important (Frehner & Lang, 2008). In these contexts, the message content tends to be more conversational and thus less structured and briefer (Cocco & Tuzzi, 2012), but are not necessarily less informative. This, coupled with the greater difficulty of typing SMS than typing at a computer, fosters the frequent sending of short SMS, which may often consist of single words. A word is defined in the current study as any string separated by spaces,

which includes isolated numerals as words, and the substitution of numbers for words. This substitution is expected in SMS communication, as writing numeric values as numbers uses fewer characters than spelling them out, i.e. "21" in comparison to "twenty one", shortening the overall length of the message. Given the proliferation of this abbreviated text speak (Power, Power, & Horstmanshof, 2007), it may be expected that responses to open ended questions given via SMS will contain significantly shorter words than other modes, having the effect of shortening the length of a message in terms of characters, but not in terms of the number of words used. Alternately, if text speak abbreviation is not used, the converse may be true. The Menzerath-Altmann law, a general principle of the structure of language, states that the longer a language construct, for example a sentence, the shorter its constituents, for example words within a sentence (Altmann, Grzybek, Naumann & Vulanovic, 2012). This would lead to the expectation that the overall shortness of an SMS should be associated with the usage of significantly longer words than other modes. Empirical evidence specific to the context of SMS used as a tool to collect research data will help to clarify which of these expectations would be met in an applied research setting.

The question remains whether a shorter response in terms of characters would necessarily be associated with less informative contents. If response lengths across modes are vastly different, it is reasonable to assume the longer responses will contain more useful information. If response lengths are similar, however, the difference in informational content may be more subtle. There is some evidence that amount of information contained in a language construct can depend on its mode of production, once the response length is controlled for. For example, oral narratives tend to be longer but contain less information than written narratives (Ravid & Berman, 2006).

Basic expectations of how much information a mode can provide are vital to informing research design in general. These issues are particularly important for a mode with inherent limitations, such as SMS. As the possibilities of SMS as a tool for data collection become apparent, and researcher interest is increasing due to its ubiquity and convenience, this is an opportune time to explore the mode's properties. The aim of this study was to ascertain the comparative response length and informative content within a single measure open-ended SMS question in the context of other modes used for psychological research. Specifically, it explored the impact of mode on open-ended response length (operationalized as number of words, average length of words, and number of characters in total) and amount of information presented within that response (operationalised as the number of points raised by participants, and the difference between height and risk-taking information given using different modes and information given at follow-up). While it expected that character limit considerations of SMS that will produce the shortest answers, it is not clear whether SMS responses will contain less information than other modes.

Method

Four hundred and sixty-three psychology students (228 female, 151 male, 84 unspecified gender) aged 16 - 55 (M = 20) answered two questionnaires during laboratory classes. An undergraduate sample was used for two primary reasons. Firstly, and most importantly, undergraduate students are one of the largest participant pools used in non-clinical psychological research (Wintre, North, & Sugar, 2001). It is therefore logical evaluations of a research mode should focus on this particular sample group.

Secondly, this allowed for the context of participation to be held consistent across all participants, regardless of conditions. Given the importance of considering context when comparing mode-based response properties (Lynn & Kaminska, 2012), all participants completed the measures in the same room, during their introductory psychology course weekly tutorials. Because the focus of the current study is the effect of the mode, rather than the role of distraction, it was important to control for environmental variables that may differ between modes, as individuals participating via SMS in a naturalistic

setting are more likely to be in public places, and thus have a more disrupted engagement with the questionnaire. Factors likely to impact on social desirability such as the presence of the researcher (Wilkerson & Martin, 2002), the task introduction and instructions were kept constant across groups, and responses were rendered anonymous by matching across questionnaires using a unique ID code drawn from personally relevant, invariant questions (e.g. the second letter of their first name).

At time 1, participants were randomly assigned to complete the first questionnaire by SMS, e-mail, online survey, or paper survey. To ensure all four modes conditions received the questionnaire at the same time, participants in the SMS and e-mail condition provided their e-mail addresses or telephone numbers, with the instructions not to check for incoming messages until the tutorial group was relocated to a computer laboratory. Regardless of mode response, all participants were moved to the same computer laboratory. E-mail questionnaires consisted of plain text e-mails sent from a custom-registered e-mail account, to be checked on provided computers. SMS used the same questionnaire sent via a commercial web-based SMS scheduling and aggregation service, to be checked on the participant's own mobile telephone. Participants in these conditions were asked to enter and submit their responses by replying to these messages. Participants in the online condition were provided with a short URL leading to the survey. Though both online and email responses were completed online, they differed in visual environment; the single large text area of email as opposed to separate text boxes for each question in the online survey. Those in the paper condition were provided with a printed version of the questionnaire on a single A4 page. Because the amount of space provided for responses can impact on response length (Dillman et al., 2009), the spaces available for the open-ended response were matched across online and paper questionnaires, though similar matching was not possible for SMS or e-mail questionnaires.

To mirror a real-life research context, the current study explored the length of responses to open-ended questions in the context of a larger questionnaire, flanked by Likert-type or single-word questions. The topic of risk taking behaviour was chosen as the content for the current study, as it is likely to be present, in varying levels, in the current sample of young adults. The current study focused only on response length as a function of mode, rather than as a function of number of questions asked. Because of this, the number of questions asked was fixed. The questionnaire consisted of five closed questions requiring short answers, and one open-ended question; (1) participant ID code (2) gender (3) age, (4) estimated height (5) the open-ended question. The questionnaire remained short to allow participants scope to provide a long answer to the open-ended question. SMS behaviour in everyday settings, as explored by Frehner and Lang (2008) was used as a guide for what might be a realistic response length via SMS. The open-ended question was formulated do that it may be reasonably answered in 280 characters or less, and wording was used to avoid participants producing closed yes/no answers. It was as follows:

> Taking a risk can include many things, such as playing dangerous sports, having a poor diet, quitting a job without another to go to, gambling, driving too quickly, or challenging a friend's opinion. Thinking about the past month, what risks have you taken?

At time 2, 2 weeks later, all participants had their height measured by the researcher, and completed a second questionnaire on paper. Paper was chosen as the comparison mode due to its long history of use in psychological research (Dillman et al., 2009). This questionnaire included a multiple choice portion, to be used as a final check of the open-ended response contents (similarly to Allison, Okun & Dutridge, 2002). Using a multiple choice, rather than open-ended question at this time was to allow a standardised comparison in risk-taking behaviours across participants. Providing a list of risk taking behaviours the participant group was most likely to have engaged in was intended to assist recall of risks

participants may otherwise not have thought of, and thus omitted from their open-ended response. Drawn from Nicholson's Risk Taking Index (Nicholson, 2005), and the 2013 State and Local Youth Risk Behaviour Survey (CDC, 2013), with items likely to be most relevant to the current undergraduate sample, as follows:

Taking a risk can include many things. Thinking about the past month, put a tick next to the risky things you have done.

- *Not wearing a seat belt (when driving, or a passenger)*
- *Riding in a car driven by someone who had been drinking alcohol*
- *Texting or e-mailing while driving a car*
- *Carrying a weapon such as a gun, knife, or club*
- *Being part of a physical fight*
- *Having sexual intercourse without using a condom*
- *Having sexual intercourse with multiple partners within a month*
- *Taken illegal drugs (including Marijuana, ecstasy, heroin or amphetamines)*
- *Taken a prescription drug (such as Adderall, Ritalin, or Xanax) without a doctor's prescription*
- *Rock-climbing*
- *Scuba diving*
- *Fast driving*
- *City cycling without a helmet*
- *Standing for election*
- *Publicly challenging a rule or decision*
- *Smoking*
- *Poor diet*
- *High alcohol consumption*
- *Quitting a job without another to go*
- *Gambling*
- *Risky investments*

Analysis and Results

At time 1, 80 participants answered via e-mail, 109 via online survey, 119 via paper, and 87 via SMS. Responses were first compared in terms of word length, formality and diversity of words used, and then their informational content (correspondence between information contained in open-ended responses, factual information, and subsequent self-report on the same topic).

Data Preparation

Prior to analysis, all variables of interest were screened for outliers (datapoints exceeding the third quartile), normality, and homogeneity of variance. There were no graphic symbols or emoticons present in responses. Though all 463 students participated in the activity, substantially fewer provided data due to absences, illegible or lost responses, and decisions not to consent to having their data used for subsequent analysis. In total, 395 participants completed the open-ended response at time 1, and 304 completed the questionnaires at both time 1 and 2. Analyses were conducted with and without outliers omitted, and with and without controlling for demographic variables of age and gender. Because conclusions were not affected they were included in final analyses.

Significant positive skewness was revealed in all variables of interest by Shapiro-Wilk and D'Agostino tests, and was corrected for by iteratively raising data to a power of .001 increments until skewness was no longer significant. Once variables were transformed to correct for skewness, screening for homoscedasticity using the Fligner-Killeen test of Homogeneity of Variances revealed nonsignificant heteroscadicity

in all variables of interest, indicating suitability of these data for ANOVA. To check that the transformation did not distort conclusions, nonparametric Kruskal-Wallis rank sum tests were carried out with the same variables as the ANOVA, using the untransformed data. These results are not reported, as they were closely in line with results from ANOVA performed on the transformed data.

Response length

Response length in characters, words, and average word length were calculated from open-ended responses at time 1. The shape of response length distribution within each mode may prove an informative measure of the length of responses that may be expected from each mode. Given the significant positive skewness found across modes, comparison of standard deviations would be insufficient to capture the dispersion of response lengths. Rather than visualise the distribution of each mode traditionally, violin plots were generated to allow clearer comparison of the distribution shape across the entirety of the distribution, including the thinner tails of the distribution (figure 1). Length of response in characters on the y axis, and the distribution shape mirrored on the x axis. Note that the positive skew of response length is shown as a bulge toward the bottom of each distribution. For clarity of comparison, the median response length for each mode is displayed on each distribution at the appropriate position on the y axis. A K-sample Anderson-Darling Test (Scholz & Stephens, 1987) was used to test whether the distributions of response length in SMS, paper, online, and e-mail data came from the same underlying distribution. The test was significant (unadjusted for ties $t = 4.086$, $p = 0.003$), indicating that the distribution of response length (in characters) was significantly different across modes. As can

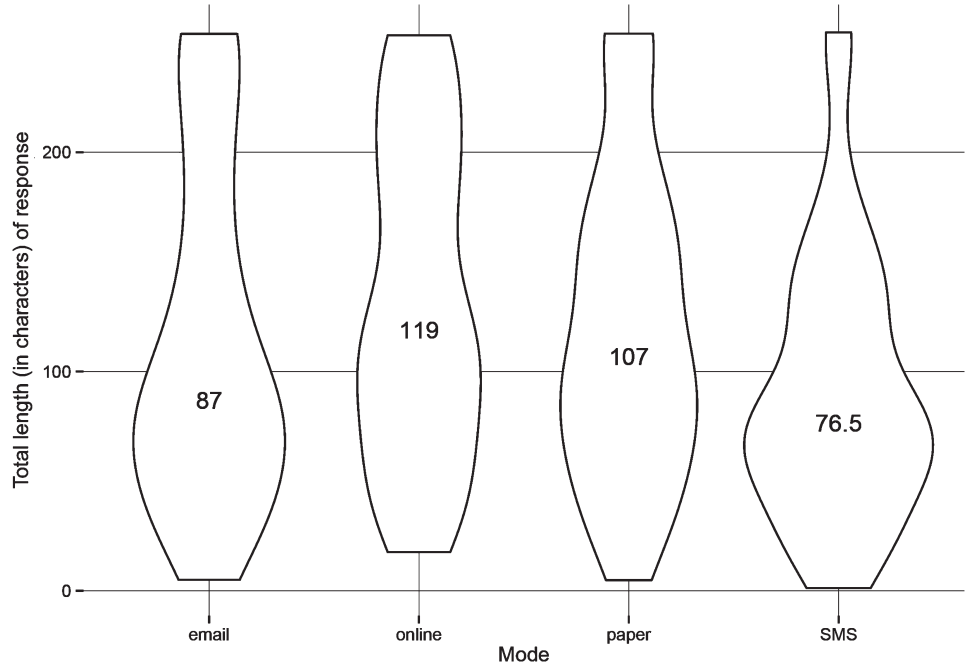

Figure 1 Distribution shape of response lengths, by mode. *Note:* The positive skew of response length is shown as a bulge toward the bottom of each distribution. For clarity of comparison, the median response length for each mode is displayed on each distribution at the appropriate position on the y axis.

Table 1 Mixed ANOVA of the impact of mode on the length of responses

Mixed ANOVA of the impact of mode on total number of characters					
	DF	SS	Mean Sq	F	Pr(>F)
Between-subjects effects					
Mode	3	13755.21	4585.07	0.84	0.4734
Residuals	300	1639431.96	5464.77		
Within-subjects effects					
Response Occasion	1	81215.01	81215.01	21.80	<.001 **
Mode x Response Occasion	3	119041.27	39680.42	10.65	<.001 **
Residuals	300	1117708.21	3725.69		
Mixed ANOVA of the impact of mode on total number words used					
	DF	SS	Mean Sq	F	Pr(>F)
Between-subjects effects					
Mode	3	190.64	63.55	0.35	0.4734
Residuals	300	53738.44	179.13		
Response Occasion	1	4033.48	4033.48	34.13	<.001 **
Mode x Response Occasion	3	5298.69	1766.23	14.94	<.001 **
Residuals	300	35456.83	118.19		
Mixed ANOVA of the impact of mode on length of words					
	DF	SS	Mean Sq	F	Pr(>F)
Between-subjects effects					
Mode	3	2.31	0.77	0.93	0.4282
Residuals	300	249.42	0.83		
Response Occasion	1	27.40	27.40	36.05	<.001 **
Mode x Response Occasion	3	9.93	3.31	4.36	0.005**
Residuals	300	228.01	0.76		

Note: **= significant at α =0.001

be seen in figure 1, SMS responses were significantly more concentrated around short response lengths than the other modes, a pattern echoed to a lesser degree by e-mail responses.

One-way between subjects ANOVA were conducted to compare the effect of measurement mode (online, e-mail, SMS or paper) on the total length of responses (in characters) at time 1 (Table 1). As expected, the effect of mode on number of characters used was small and significant, though the effect size was very small $F(3,381) = 4.91$, $p = .002$, partial $\eta^2 = 0.04$. Comparing across modes, SMS had the fewest characters, followed by e-mail, and paper, while online had the most characters (Table 2). Mode was responsible for 4% of the variance in the number of characters used in responses. Post hoc comparison by way of Tukey's HSD indicated that SMS responses were significantly shorter than online ($p = .001$) and paper ($p = .031$) responses.

The effect of mode on the average length of words used in responses was small and significant, $F(3,381) = 3.4$, $p = .018$, partial $\eta^2 = 0.03$. Comparing across modes, email had the longest words on average, followed by SMS, online surveys, and paper. Results supported the Menzareth-Altmann law's assertion that longer language constructs are associated with shorter constituent elements, as the modes with the longest open-ended responses (paper and online) had shorter words on average. The smallness of the effect makes sense given that word lengths and text lengths were quite similar (Grzybek,Stadlober, & Emmerich, 2007). Mode was responsible for 3% of the variance in the length of words used in responses.

Table 2 Response lengths at time 1

	Email	Online	Paper	SMS
Average number of characters	113.68	130.84	120.06	91.95
Average number of words	18.42	23.13	21.14	14.8
Average length of words (in characters)	5.36	4.96	4.94	5.24
Count of simplifications and substitutions	5	8	7	5
Count of reductions	8	7	15	5
Count of spelling errors	8	2	2	2
Count of typographical errors	1	5	6	0
Average token-type ratio	0.94	0.92	0.93	.095
Average Yule's K	60.71	83	77.71	56.71

Note: This summary is drawn from all responses at time 1, $n = 385$.

Post-hoc comparison by way of Tukey's HSD indicated that paper responses had significantly shorter words than email responses ($p = 0.04$).

The effect of mode on the average number of words used in responses was pragmatically small (on the order of five words), but significant, $F(3,381) = 7.46$, $p < .001$, partial $\eta^2 = 0.06$. Comparing across modes, SMS had the fewest words, followed by email, and paper, while online had the most words (Table 2). Mode was responsible for 6% of the variance in the number of words used in responses. As with the effect of mode on number of characters used, Post-hoc comparison by way of Tukey's HSD indicated that SMS responses had significantly fewer words than online ($p < .001$) and paper ($p = .003$) responses. Taken together, these results suggest that SMS responses were shorter overall due to the use of fewer words than other modes, rather than the use of shorter words.

Formality and diversity of words used

Two independent coders categorised spelling modifications in open-ended responses, following Combes, Volckaert-Legrier, & Largy (2012)'s typology; occurrences of substitutions and simplifications (i.e. "I'll" instead of "I will," but not possessives like "friend's"), reductions (i.e. "mess" rather than "message"), spelling errors (i.e. "nessecarry" rather than "necessary"), and typos (i.e. "hjt" rather than "hit"), see Table 2. Chi-square contingency tables were revealed that there was no significant differences between modes in the presence of substitutions and simplifications, spelling errors, or typos ($\chi^2(3) = 4.79$, $p = 0.95$; $\chi^2(3) = 3.44$,$p = 0.32$; $\chi^2(3) = 0.2$, $p = 0.97$ and $\chi^2(3) = 6.15$, $p = 0.104$ respectively).

The textual content of the open-ended responses can also be thought of in terms of the variety in the vocabulary used. The more frequently the same words are used, the more limited the vocabulary (Köhler, Altmann & Piotrowski, 2005). This can be measured in terms of the vocabulary size, the number of unique words used in a response, or the token-type ratio, which is the weighted range of vocabulary for number of words in the text being analysed. However, the token-type ratio is highly dependent on the length of text, and given the shortness of responses in the current study (averaging less than 150 characters) is therefore a potentially flawed measure of lexical richness (Tweedie & Baayen, 1998; Popescu, 2009). We will therefore also examine the text in terms of Yule's K, which is a length-independent measure of the diversity of the vocabulary (Miranda-Garcia & Calle-Martin, 2006).

Both the token-type ratio and Yule's K were calculated for each open-ended response. As can be seen in Table 2, both statistics were similar across modes. ANOVA confirmed a lack of significant

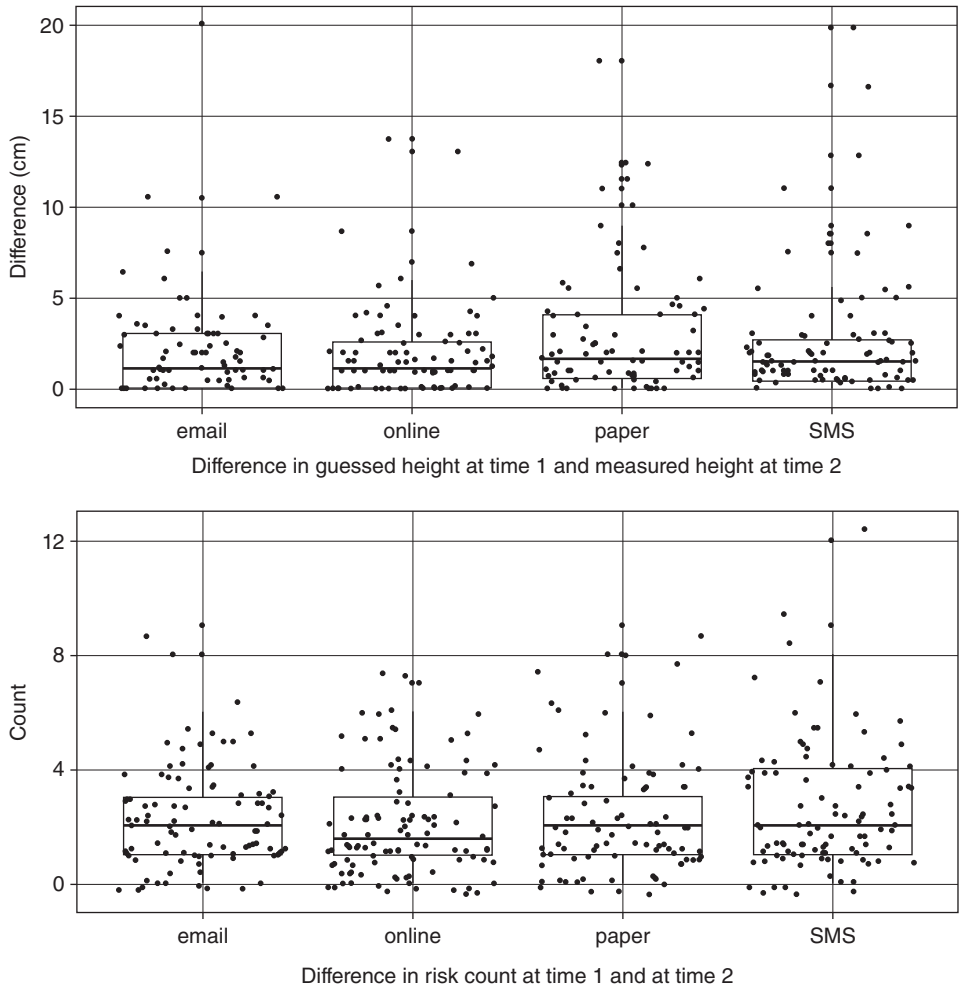

Figure 2 Differences in guessed and actual height (above) and count of self-reported risk-taking behaviours (below), split by mode. The points have been jittered on the horizontal axis to avoid overplotting, thus giving a more readable view of the distribution of points on the vertical axis. *Note:* The vertical axis of the upper plot has been truncated at 20cm for readability. There are 4 outliers in email condition, and 4 in online conditions extending up to 70cm.

difference each of these statistics; neither token-type ratio nor Yule's K significantly differed by mode ($F(3,216) = 1.637$, $p = .182$, partial $\eta^2 = .02$ and $F(3,216) = 1.018$, $p = .385$, partial $\eta^2 = .014$ respectively). This suggests that mode did not impact specifically on the diversity of the vocabulary used.

Informational Content

The effect of mode on the accuracy of responses as measured by the difference between reported height at time 1, and measured height at time 2 (shown in figure 2) was not significant, $F(3,358) = 0.274$, $p = .844$, partial $\eta^2 = 0.02$, indicating that mode did not impact on response validity as operationalised as correctness of self-reported height.

Descriptively, participants reported an average of approximately three risks reported in open-ended responses overall. This did not significantly differ by mode, $F(3,381) = .571$, $p = .634$, partial $\eta^2 = 0.03$, with participants responding by paper having an average of 3.4 risks, by e-mail 3.3 risks, online 3.2 risks, and by SMS 3 risks. This suggests that mode did not have an impact on the amount of informative content within responses.

The absolute difference between number of risks mentioned at time 1, and number of risks reported at time 2, and the number of risks reported in the 'other' category at time 1 were used as a measure of the comprehensiveness of open-ended responses. Comparison in self-reported number of risks across time points is only to be interpreted as a broad check of consistency between self-reported risk at time 1, and risks selected at time 2. The two measures may not directly correspond due to participants engaging in risky behaviours between measures, or because the multiple choice list at time 2 acted as a reminder of risks forgotten at time 1. Additionally, they do not provide an absolute measure of how similar the two sets of risks are. However, this comparison is meaningful in that it helps to disentangle whether few risks reported at time 1 are due to a genuine lack of risk taking, or a poor response. There was no significant difference in the number of 'other' risks identified by participants at time 1, $F(3,319) = .297$, $p = .83$, indicating that random assignment across modes was successful, and that subsequent analyses were truly disentangling mode effects from individual differences in risk taking.

Though the number of risky behaviours reported in the open-ended question categorised as "other" indicate that the specific risks used in the questionnaire at time 2 were not a comprehensive reflection of the risks taken by the current sample, there was not a significant effect of mode on the difference between number of self-reported risks reported at time 1 and time 2, $F(3,379) = 1.82$, $p = .124$. This indicates that random assignment across modes was successful in ensuring a relatively even distribution of low and high risk takers responding in each mode, supporting the assertion that mode did not affect the amount of informative content within responses.

Discussion

Results consistently demonstrated that the responses to an open-ended question embedded in a larger questionnaire is statistically, but not pragmatically, impacted upon by mode. In line with previous research (i.e. Cocco & Tuzzi, 2012) SMS responses were significantly shorter in total length, having fewer characters and fewer words than other modes. However, SMS here did not contain more incorrect information, or fewer points of information, than other modes. At an average of 92 characters, SMS response lengths to the open-ended question were marginally shorter than the average SMS length reported in Frehner and Lang (2008)'s corpus. This makes sense, as the corpus looked at the total length of SMS, while the current study looked only at the length of the open-ended question, a portion of the total SMS response.

In accordance with the Menzareth-Altmann law, the shorter overall total length in terms of characters of SMS responses was reflected in longer average word lengths than the other three modes. Both online and paper responses had the opposite tendency of longer responses overall, but shorter constituent words. It is unlikely that the 160-character limit of SMS is driving this use of fewer words, as all four modes had an average response length of less than 160 characters. As SMS responses did not contain significantly more substitutions or reductions than the other modes, neither is it due to the use of numeric phonological substitution discussed in Power et al. (2007). Given that mode did not impact on the token-type ratio, it does not appear that SMS has fewer words due to less redundancy.

Though responses were longer than those yielded by SMS, the pattern of response lengths in terms of characters, number of words, and word length for e-mail was far more similar to SMS than online or paper responses. Considering the shape of the distribution of response length in characters, SMS and

e-mail had a similar shape of clearly positively skewed distributions bulging about median responses, though the bulge was more extreme for SMS, while paper and online responses had far more diffuse, almost bimodal response length distributions. Cocco and Tuzzi (2012)'s allusion to SMS engendering "virtual presence" of the researcher may be considered and applied here, as the similarity of SMS and e-mail responses in the current study may be due to the social aspect of use of those modes of communication. SMS and e-mail may be conceptually grouped as modes typically used for bidirectional social communication, while paper and online responses are unidirectional modes generally not encountered in social circumstances. SMS and e-mail may share the social rather than informational motive for use, where the fact the message was sent is more important than its content (Frehner & Lang, 2008), and the message content tends to be briefer (Cocco & Tuzzi, 2012), even if informal language such as text speak was not used. The lack of textspeak in SMS responses in the current study may be an artefact of the formality of responding in a teaching laboratory context.

As the first step in examining the impact of SMS as a data collection mode on response length and informational content, this study made a number of somewhat arbitrary choices. Firstly, the operationalization of informational content was not ideal. Using a directly observable phenomenon in parallel with a less visible psychological construct (risk taking) was valid, but the choice of height was problematically easy, reflected in the very small effect size revealed in analyses. Similarly, attempting to corroborate robustness of an open-ended question with a more standardised question format is valid. However, the correspondence between an open-ended question, and later multiple-choice question is not ideal. There are individual differences in response behaviour to particular question types, for example culturally-based polarisation in Likert responses (Heine, Lehman, Peng, & Greenholtz, 2002). Further, the time between measures may have allowed more risk taking to occur. Given the random assignment of participants to different mode response conditions, it is likely that these problems would have added noise to the data rather than systematically bias conclusions. This noise may have contributed to mode explaining only a small amount of variance in response length. Additional unmeasured variables that impact on verbosity, such as personality traits (Mairesse, Walker, Mehl, & Moore, 2007), or verbal fluency (Gold & Arbuckle, 1995), may have added further noise to the data.

The choice to conduct the current study in a teaching laboratory context is both a strength and a weakness. Environmental effects such as being in a hurry when receiving SMS, having to search for an e-mail in an overfull inbox, having to configure a browser to correctly load an online survey, and having to find a lost paper survey were controlled for by having all participants complete the questionnaire in a standardized environment and timeframe. This yielded a clearer comparison of the impact of mode alone on response length and informative content, in particular controlling for time pressure, but removes opportunity to examine whether time pressure, the formality of the response situation, or other environmental factors may further modify the response length to research conducted via SMS in a real-world research setting. A related concern is that participants in the email condition were constrained to responding using computers, whereas in everyday life they may respond to e-mails using their smartphones. This highlights the importance of considering the amount of space provided for answers by response mode, as the ability to use smart phones to respond to both e-mails and SMS may make the two modes more similar in a real-world setting than they were in the current study. This limits the ability to draw clear conclusions about e-mail as a mode.

Acquiescence, resulting in demand effects, is important to consider when exploring the impact of mode on response properties (Kelly et al., 2008). Participants were aware from the outset that the focus of the researcher was the usefulness of SMS as a tool for research, and that across modes their responses would be screened for a number of indicators of response quality. Though response length and amount of information conveyed was not explicitly mentioned, participants may have provided longer or more

detailed responses due to perceived demand characteristics than they otherwise would have offered, particularly in the SMS condition given the focus of the researcher.

Because the focus was on cross-modal comparison, this study explored response length with a relatively arbitrary set of five questions. Doing so limits generalizability, but opens avenues for further investigation seeking to focus specifically on SMS. In terms of the SMS character limit, there was scope for the researcher to send more questions. The relatively short responses demonstrated scope for participants to provide more answers. Pragmatically speaking, the more questions there are, the less space there is for participants to respond to each. The nature of questions also come into play, as more complex questions require more complex explanations (thus limiting how many the researcher may send), and more detailed answers (thus limiting how long responses will be). It is therefore likely that a different number of questions, or different configuration of question types (i.e. Changing a Likert-style response to an open-ended question) may impact on response lengths overall. Future research could vary the number of questions to uncover the impact of question complexity on response length, and also whether more questions lead to longer responses overall, or truncated answers to each question. It could also somewhat overcome the limitations of the current study by choosing an observable metric other than participant height, and using a single instrument multiple times, rather than two different measures of the same construct as was done in the current study.

This paper applied both nonparametric statistics and transformation with parametric statistics to deal with the zero-bounded highly skewed variables, and it should be noted that p values were very clearly nonsignificant. However, it remains that the conclusions reached are largely based on "proving the null," in a context of very small effect sizes. Future researchers should certainly employ statistics better suited an expectation of null hypothesis confirmation, such as taking a Bayesian approach.

This study compared response length and informative content for an open-ended SMS question to other modes of data collection used for psychological research. SMS response lengths were statistically significantly shorter than those yielded in other modes, but SMS responses did not contain less valid information, or less information overall. This suggests that, for a question requiring only a short response, SMS is comparable to other modes as a viable mode for research involving open-ended responses embedded in a questionnaire.

References

ACMA. (2011). *ACMA Communications report 2010–2011*.

Allison, L. D., Okun, M. a., & Dutridge, K. S. (2002). Assessing volunteer motives: A comparison of an open-ended probe and Likert rating scales. *Journal of Community and Applied Social Psychology*, 12, 243–255. doi:10.1002/casp.677

Altmann, P. Grzybek, S. Naumann, R. Vulanović (2012, eds), *Synergetic linguistics: Text and language as Dynamic Systems*, Wien: Praesens Verlag.

Anhoj, J., & Moldrup, C. (2009). Feasibility of collecting diary data from asthma patients through mobile phones and SMS (short message service): response rate analysis and focus group evaluation from a pilot study. *Journal of medical internet research*, 6(4), e42. doi:10.2196/jmir.6.4.e42

Battestini, A., Setlur, V., & Sohn, T. (2010). A large scale study of text messaging use. *12th International Conference on Human Computer Interaction With Mobile Devices and Services* (pp. 1–10). New York: ACM. doi:10.1145/1851600.1851638

Cocco, M., & Tuzzi, A. (2012). New data collection modes for surveys: A comparative analysis of the influence of survey mode on question-wording effects. *Quality and Quantity*. doi:10.1007/s11135-012-9708-1

Combes, C., Volckaert-Legrier, O., & Largy, P. (2012). Automatic or controlled writing? The effect of a dual task on SMS writing in novice and expert adolescents. *Lingvisticae Investigationes*, 35(2), 199 – 217.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2009). Internet, Mail, and Mixed-Mode Surveys *(Third.)*. New Jersey: John Wiley & Sons.

Doring, N. (2002). "Kurzm wird gesendet" - Abkurzungen und akronyme in der SMS-Kommunikation, muttersprache. *Vierteljahresschrift fur Deutsche Sprache*, 112(2), 07 – 114.

Frehner, C., & Lang, P. (2008). *Email - SMS - MMS: The linguistic creativity of asynchronous discourse in the new media age*. Bern, Germany: International Academic Publishers.

Gold, D. P., & Arbuckle, T. Y. (1995). A longitudinal study of off-target verbosity. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 50(6), P307 – P315.

Grzybek, P., Stadlober, E., & Kelih, E. (2007). The relationship of word length and sentence length: the inter-textual perspective. In *Advances in Data Analysis,* (pp. 611 – 618). Springer Berlin Heidelberg.

Haller, D., Sanci, L., Sawyer, S., Coffey, C., & Patton, G. (2006). R U OK 2 TXT 4 RESEARCH? Feasibility of text message communication in primary care research. *Australian Family Physician*, 35(3), 175 – 176. Retrieved from http://www.racgp.org.au/afp

Heine, S. J., Lehman, D. R., Peng, K., & Greenholtz, J. (2002). What's wrong with cross-cultural comparisons of subjective Likert scales? The reference-group effect. *Journal of personality and social psychology*, 82(6), 903. doi: 0022-3514.82.6.903

Johansen, B., & Wedderkopp, N. (2010). Comparison between data obtained through real-time data capture by SMS and a retrospective telephone interview. *Chiropractic and osteopathy*, 18(10), 1 – 7. doi:10.1186/1746-1340-18-10

Kelly, D., Harper, D. J., & Landau, B. (2008). Questionnaire mode effects in interactive information retrieval experiments. *Information Processing and Management*, 44(1), 122 – 141. doi:10.1016/j.ipm.2007.02.007

Kelly, W. E. (2002). Worry and sleep length revisited: Worry, sleep length, and sleep disturbance ascribed to worry. *The Journal of Genetic Psychology*, 163(3), 296 – 304. doi:10.1080/00221320209598685

Köhler, R., Altmann, G., Piotrowski, R.G. (2005), *Quantitative linguistics: An international handbook.* Berlin: Mouten de Gruyter

Kuntsche, E., & Robert, B. (2009). Short Message Service (SMS) technology in alcohol research: A feasibility study. *Alcohol and Alcoholism*, 44(4), 423 – 428. doi:10.1093/alcalc/agp033

Lynn, P., & Kaminska, O. (2012). The impact of mobile phones on survey measurement error. *Public Opinion Quarterly*. doi:10.1093/poq/nfs046

Mairesse, F., Walker, M. A., Mehl, M. R., & Moore, R. K. (2007). Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of artificial intelligence research*, 457 – 500.

Mackay, M. M., & Weidlich, O. (2009). Austrailan Mobile Phone lifestyle index. *Specialist*. Australian Interactive Media Industry Association Mobile Industry Group. Retrieved from http://www.aimia.com.au/enews/mobile/090929 AIMIA_Report_FINAL.pdf

Miranda-García, a., & Calle-MartÍn, J. (2006). Yule's Characteristic K revisited. *Language Resources and Evaluation,* 39(4), 287 – 294. doi:10.1007/s10579-005-8622-8

Polite, J. R. (2001). Das Handy als "personliches Medium" zur aneignung des Short Message Service (SMS) durch Jugendliche. *kommunikation@gesellschaft* (Vol. 2, pp. 1 – 19).

Popescu, I. I. (2009). *Word frequency studies* (Vol. 64). Walter de Gruyter.

Power, M. R., Power, D. J., & Horstmanshof, L. (2007). Deaf people communicating via SMS, TTY, relay service, fax, and computers in Australia. *Journal of deaf studies and deaf education*, 12(1), 80–92. doi:10.1093/deafed/enl016

Ravid, D., & Berman, R. (2006). Information density in the development of spoken and written narratives in English and Hebrew. *Discourse Processes,* 41(2), 37–41. doi:10.1207/s15326950dp4102

Steeh, C., Buskirk, T. D., & Callegaro, M. (2007). Using text messages in U.S. mobile phone surveys. *Field Methods*, 19(1), 59–75. doi:10.1177/1525822X06292852

Tomlinson, M., Solomon, W., Singh, Y., Doherty, T., Chopra, M., Ijumba, P., Tsai, A. C., et al. (2009). The use of mobile phones as a data collection tool : A report from a household survey in South Africa. *BMC Medical Informatics and Decision Making*, 8(1), 1–8. doi:10.1186/1472-6947-9-51

Tweedie, F., & Baayen, R. (1998). How variable may a constant be? Measures of lexical richness in perspective. *Computers and the Humanities,*32(5) 323–352.

Wilkerson, M., & Martin, L. (2002). Socially desirable responding in computerized questionnaires : When questionnaire purpose matters more than the mode. *Journal of applied social psychology*, 32(3), 544–559.

Wintre, M. G., North, C., & Sugar, L. A. (2001). Psychologists' response to criticisms about research based on undergraduate participants : a developmental perspective. *Canadian Psychology*, 42(3), 216–225. doi:2001-11323-005

About the Authors

Erin Walsh is a post doctoral research fellow at the Centre for Research on Ageing, Health and Wellbeing. Alongside research, she works as a scientific illustrator, offering assistance with stimuli preparation, data visualization, and figure preparation for scholarly publication or engaging the general public with science.
Address: Centre for Research on Ageing, Health and Wellbeing, Research School of Population Health, Australian National University, 54 Mills Road, Acton, ACT, 2601.

Jay Brinker is both an academic and a clinician who spends most of her time conducting research, teaching and supervising clinical and research students at Swinburne University.
Address: Psychological Sciences, Swinburne University of Technology, Hawthorn, Vic 3122.