As you Likert – cross-mode equivalence of administering lengthy self-report instruments via text message

Walsh, E.I, & Brinker, J.

This paper has been accepted in the peer-reviewed 2014 ACSPRI[1] Social Science Methodology conference proceedings. It is presented as a word document because the formatted conference proceedings layout was not available at time of submission.

This paper was presented at the ACSPRI Social Science Methodology Conference, December 2014, Sydney, Australia.

One of the most widely used data services worldwide, Short Messaging Service (SMS) offers an unprecedented opportunity for researchers to communicate with participants at any location, or time. One concern when using SMS for research is whether the mode's brevity may make it unsuitable for administering multi-question, self-report psychological instruments originally developed for paper or online administration. Across two studies, this paper explores the psychometric properties and cross-mode measurement invariance of self-report, likert-style psychology questionnaires administered via SMS. The first study (*n*=417) examined this using different length variants of the same instrument, while the second (*n*=911) used instruments of varying lengths. Results demonstrated that, whilst some questionnaires were problematic, a self-report likert-style instrument as long as sixty items can be administered by SMS, with comparable response rates, internal reliability, and factor structures to online administration. However, in instruments over ten items in length, mean responses tended to be higher, leading to lack of equivalence in terms of latent means and intercepts.

[1] Australian Consortium of Social Science and Political Research Incorporated

A measure is invariant if individuals with the same standing on a construct receive the same score on an instrument intended to measure that construct (Schmitt and Kuljanin, 2008). Undesirable variance can occur where the same instrument is given to populations who engage with it in a different way (e.g. Wicherts, Dolan, and Hessen, 2005), or where the researcher seeks to compare scores from different versions of an instrument administered to different groups. This can happen when participants respond to researchers using multiple modes, such as web, online or email self-report surveys being used in parallel (Dillman, Smyth, & Christian, 2009). Recently, cross-modal invariance is becoming important as mixed-mode research becomes more common (De Beuckelaer & Lievens, 2009), and new modes offer unique research opportunities (Cocco & Tuzzi, 2012). One such new mode is short message service (SMS), a ubiquitous text-based functionality of mobile telephones. One of the most widely used data services worldwide (Kuntsche & Robert, 2009), most Australians use SMS daily (Mackay & Weidlich, 2009). This is an unprecedented opportunity for researchers to communicate with participants at any location, or time (Haller, Sanci, Sawyer, Coffey, & Patton, 2006). The practicality of using SMS for research has been improved by the emergence of online bulk services that allow researchers to cheaply schedule SMS in advance to an arbitrary number of participants (Steeh, Buskirk, & Callegaro, 2007). Exploration of the cross-modal validity of using SMS as a tool for research can build upon techniques used and issues raised by previous cross-mode research.

Cross-modal measurement invariance research peaked during the transition from paper to online-based administration of questionnaires, commonly applying multi-group confirmatory factor analysis to compare the equivalence of instruments administered via different modes (e.g. Leung & Kember, 2005; Richardson & Johnson, 2009; Schmitt & Kuljanin, 2008; Vecchione et al., 2012). Web and paper administration of the same questionnaires have been found to be equivalent across many topics, including in-class

teaching feedback (Leung & Kember, 2005), depression (Yu & Yu, 2007), perceived stress and depressive thinking (Herrero & Meneses, 2006), tobacco dependence (Richardson & Johnson, 2009), and organisational engagement (De Beuckelaer & Lievens, 2009). Few investigations found differences in terms of underlying factor structures of paper versus online administration, but this has been observed (e.g. Hirai, Vernon, Clum, and Skidmore, 2011). The most commonly reported effect of mode on measurement equivalence relates to mean responses, with web questionnaires tending to have systematically higher scores in general (Vecchione et al., 2012), and subsequently higher latent mean scores (Cole, 2006; Meade, Michels, & Lautenschlager, 2007).

To date, the possibilities of administering relatively lengthy self-report instruments via SMS have not been explored. Though there are some examples of single administrations of longer instruments via SMS (e.g. Lee et al., 2013), SMS tends to be used for brief self-report measures specifically designed for repeated administrations, typically the same few likert-style questions per sampling occasion. This is likely because of the inherent brevity of SMS as a communication tool. The 160 character-per-message has historically restricted how much information could be provided in an SMS. As SMS technology develops, the 160 character per message limit has been bypassed by support for stitching multiple SMS together. This allows for much longer SMS communications. The longest instruments administered via SMS in the literature currently stand at 23 (De Lepper et al., 2013) and 24 (Lee et al., 2013) items. There is evidence from research with other modes that the length of a questionnaire can meaningfully impact participant engagement (Dillman et al., 2009) and in turn, data quality and response rates (e.g. Burchell & Marsh, 1992; Jepson et al. 2005; LaMar & Gale, 1982). One factor in this is whether the instrument is presented all at once, or broken across mutliple pages or screens (Cocco & Tuzzi, 2012). Cross-modal comparisons involving SMS use instruments too brief to explore these issues, because there are too few items to be

amenable to a factor analysis. Instead, such investigations typically focus on response rates, participant feedback, and use descriptive statistics for comparison.

Repeated SMS measures have correlated with prospective online baseline questionnaires (Lim, Sacks-Davis, Aitken, Hocking, & Hellard, 2010), and retrospective telephone interview self-report measures (Johansen & Wedderkopp, 2010; Whitford et al., 2012). A large scale survey of influenza vaccination coverage found comparable odds ratios of vaccination within those surveyed via telephone interview or SMS (Bexelius et al., 2009). Physical activity diaries completed via online, paper and SMS diaries were similar in terms of mean physical activity reported (Lagerros, Sandin, Bexelius, Litton, & Löf, 2012; Shapiro et al., 2008). Comparison between SMS and app administration of a serious mental illness assessment revealed similarly congruent results in terms of descriptive statistics (Ainsworth et al., 2013). Though this is promising, there are several reasons to suspect that SMS responses may not be equivalent to online or paper counterparts when administering pre-existing questionnaires.

The visual presentation of a self-report instrument is important (Dillman et al., 2009; Richardson & Johnson, 2009), impacting response rates (Jansen, 1985) by affecting how difficult it is to read questions (Smith, 1993), and the length of a response (Fuchs, 2009). Due to the size of mobile telephone screens, questions sent via SMS will almost invariably be presented in a smaller, more cramped format than those presented online, or by paper. This may lead to participants missing items when responding via SMS, particularly for longer instruments where the SMS received will be particularly dense. The way in which a response is recorded also differs across modes in a way that may impact upon response rates – the experience of typing on a QWERTY smartphone touchscreen is more clumsy than using a full-sized computer keyboard (Page, 2013), or writing with a pen. This may lead to mistakes

that threaten the internal reliability of an instrument, such as entering a ‘5’ when the next along,‘4’, was intended.

The greater difficulty associated with reading and responding to questions, the greater the participant burden, which in turn impacts on response rates and attrition (Bolger, Davis, & Rafaeli, 2003). Where participants anticipate too high a burden, they are likely to drop out. Comparative response rates across modes are an important consideration (Leung & Kember, 2005). Problems in recruitment, or pre-inclusion attrition (where participants drop out between recruitment and active particiaption) can indicate problems and limits of a methodology (Cook & Campbell, 1979). There is little point establishing the validity of a mode if participants are unwilling to use it, particularly if it is being compared to other modes with established, generally accepted, response rates (e.g. self-report research carried out on line has an estimated approximate 40% response rates; Cook, Heath, & Thompson, 2000; Shih, 2008).

One of the greatest strengths of SMS as a research mode – the ability to contact participants regardless of their location – may also be a weakness, in that the researcher has no control over what distractions may be present as a participant responds. A noisy environment can disrupt attention (Banbury & Macken, 2001), and may lead to disengagement and perhaps even premature submission of an incomplete response. This would scale with the length of the instrument, as the greater the number of items, the more opportunity there is for missingness. Participants are very unlikely to complete a paper or online instrument in noisy settings, such as on public transport or at the pub, but may well attempt an SMS response in those locations. This could lead to greater missingness of SMS responses in comparison to other modes.

The current state of SMS technology, and the formative state of the literature relating to the measurement invariance of using SMS to administer pre-existing psychological

instruments, begets two interrelated questions. Firstly, to what extent does administration via SMS impact upon the psychometric properties of pre-existing psychological instruments? Secondly, how many questions may be asked via SMS before response rates and psychometric properties are compromised? Over two studies, this paper explores cross-modal invariance of Likert-style instruments of varying lengths administered by SMS in comparison with online and paper administrations.

Analytical approach

There are many different ways to analyse measurement invariance (Borsboom, 2006). The current approach combined statistics used in the literature to evaluate the quality of SMS data, and the factor-analytic technqiues used elsewhere in the cross-mode invariance literature. Comparison began by looking at response rates (RR), and usable response rates (UR; i.e. completing all constituent items with responses within the appropriate ranges). Scored instrument means were compared by Welch two sample t-tests. As the distribution of scores are also informative (Herrero & Meneses, 2006; Hirai et al., 2011), Anderson-Darling k sample tests (with 1000 simulated replicates) were used to establish whether the score distributions were significantly different. Internal reliability was evaluated by Chronbach's alpha (as in Herrero & Meneses, 2006; Hirai et al., 2011; Leung & Kember, 2005).

As recommended by Leung and Kember (2005), feasability of results was established by comparison with factor structures reported in the literature. The factor structure suggested by visual inspection of scree plots and variable factor maps for each mode was evaluated by model fitting and inspection of χ^2 tests of whether the number of factors specified were sufficient. Following the recommendations established by Vandenberg and Lance's (2000) synthesis of the measurement invariance literature, multi-group confirmatory factor analysis was then carried out, primarily using the semTools package in R (Pornprasertmanit, Miller,

Schoemann & Rosseel, 2013). First, a baseline model with the same factor structure for both groups is fit to the data. This test of *configural invariance* evaluates whether the same factor structure is appropriate for both groups. Subsequent models are compared with this model. Measurement invariance is supported if the models do not significantly differ in fit from the configural invariance model (denoted by a non-significant $\delta\chi^2$), but do significantly fit the data (indicated by significant χ^2). *Metric invariance* (also known as 'weak invariance') tests whether the factor loadings are equivalent, by constraining them to be the same across groups. *Scalar invariance* tests whether the factor loadings and intercepts are equivalent by constraining them to be the same across groups. If metric and scalar invariance are present, a final test, *general factor invariance,* constrains factor loadings, intercepts and means to be the same across groups – a significant χ^2, and non-significant $\delta\chi^2$ for this test provides strong evidence for equivalence.

Study 1

The aim of this study was to investigate the cross-modal measurement invariance of administering instruments of varying lengths via SMS, as opposed to paper. To this end, short forms of the Ruminative Thought Styles questionnaire (RTS; Brinker & Dozios, 2009) were constructed.

Participants

Participants were 417 undergraduate students aged 17-66 (M=22), 53% female. 42 were assigned to complete the five item RTS (20 via paper, 21via SMS), 46 the ten item version (20 via paper, 26via SMS), 46 the fifteen item version (20 via paper, 26 via SMS), and 283 the original twenty item RTS (120 via paper, 163 via SMS). Participants responding via SMS were recruited by way of posters. Those responding via paper were recruited in person upon completion of other experiments, or at a table outside lecture halls. All participants were offered thirty minutes course credit as an incentive, regardless of the version or the mode used to respond.

Materials

The RTS (Brinker & Dozios, 2009) is a self-report rumination questionnaire with 20 items loading on a single factor. Responses are given on a likert scale of 1 through 7. Three short form versions (fifteen item, ten item, and five item) of the RTS were created for the purposes of this study. The short forms were created by pooling RTS data from eight past studies with university undergraduate samples. Following analyses confirming homogeneity of RTS and demographics across samples (mean participant age=26, *SD*=16.66, 34% male), the short form versions were constructed by iteratively dropping the items to maximise Chronbach α

and single-factor loadings. The resultant short forms had good internal consistency, with α ranging from .86 through to .93.

The original RTS instructions direct respondents to circle the correct number corresponding to each item, which is untenable via SMS. An alternate method of responding amenable to SMS completion was developed:

> *"For each of the RTS items, rate how well the item describes you, with 1 indicating not at all, through to 7 indicating very well. Indicate which question you are responding to, a full stop, and then your answer. Separate each question with the letter x. Answer all of the questions in order. Example: "1.3x2.5x3.5" etc."*

To control the differences in response experience across modes, the short form versions of the RTS administered by paper were given the same response instructions as those administered by SMS. So, when responding by paper, participants wrote their answers in the same format as provided in the example above. To establish whether this change of instructions affected responses, some (*n*=20) participants completed the twenty item RTS with the new instructions, while the rest (*n*=100) completed it using the original RTS instructions. All participants using the new instructions were asked to rate the difficulty of writing their responses on a scale of 0 (very difficult) to 5 (very easy).

Design and Procedure

This pseudo-experiment used a 2 (modes) x 4 (length versions) design. For practical reasons, response mode was not randomly assigned. This was because the physical presence of students on campus at the time of completing the questionnaire was required for paper completions (as participants had to obtain the survey, complete it, and then return it in person), but not SMS completions (as participants could provide their contact details and then

leave). Recruitment for SMS and paper versions was carried out simultaneously, but separately. The two response modes were advertised as though they were separate and mutually exclusive studies, with participants effectively self-selecting which mode they would use depending on which study they signed up to. Assignment to length version group was random.

Those responding via SMS indicated their consent by texting "yes" and a time and date they preferred to respond. Instructions for completing the RTS were sent at the designated time, followed by the items fifteen minutes later. Due to the length constraints of SMS, the instrument was sent in five item chunks scheduled to arrive at the same time (i.e. the twenty item version was split across four messages). Demographic and feedback questions were sent the day after the questionnaire. Those responding via paper collected the RTS from the researcher, returning it anonymously by way of a submissions box.

Results

Response rates were relatively high and consistent across the ten, fifteen and twenty item versions, with no clear trend in terms of item missingness diminishing usable responses as the number of items increased (Table 1). While the five item version demonstrated a much higher response rate and usable response rate than the other length versions, it did not demonstrate cross-mode equivalence. SMS and paper administrations of the five item version differed significantly in terms of means, distributions, and underlying factor structure (Table 1, Table 2), precluding further multi-group factor analysis. Inspection of the communalities, the proportion of variation in the items explained by the factors specified, can indicate whether a particular item is responsible for poor model fit. Ideally, communalities should approach 100 (indicating all variance in a given item is explained by the factor). Different items were problematic in the SMS and paper responses; item 5 was the most problematic in SMS with a

commonality of just 2, whilst item 3 was most problematic in paper with a commonality of just 10. Post-hoc analysis of participants completing the five item version found that those completing the instrument via paper were significantly younger than those via SMS (paper mean age = 21, SMS = 34, $t(20)=4.101$, $p<.001$).

Table 1. Missingness, descriptive and internal consistency comparisons between modes

RTS			SMS			Paper			*t*	*p* K
Version	RR	UR	M	SD	α	M	SD	α		
5	99%	90%	21.25	5.6	.79	24.56	3.55	.67	2.19, *p*=.03	.029
10	73%	57%	47.32	10	.68	42.67	8.55	.89	1.41, *p*=.17	.397
15	76%	73%	66	16	.88	60	15	.90	1.25, *p*=.21	.193
20	73%	67%	84.81	20	.91	88.45	17.82	.90	1.45, p=.15	.22

Note. RR = response rate, the percentage of participants who at least partially completed the SMS instrument. UR = usable response rate, the percentage of participants who correctly completed all items of the SMS instrument. A *p*K value less than 0.05 signifies a significant result for the Anderson-Darling k-sample test, indicating the scale's total scores come from different underlying distributions depending on mode.

The longer ten, fifteen and twenty item versions of the RTS had consistently excellent internal consistency comparable with online completions ($\alpha>.85$). Reasonable SMS response rates ($\approx$73%) diminished somewhat by item missingness within responses, particularly due to

out-of range responses. Once scored, there were no significant mean or distribution differences between SMS and paper administration of the ten, fifteen, or twenty item versions of the RTS (Table 1). Multi-group factor analysis confirmed configural, metric, scalar and general factor invariance across modes of administration for all length versions but the shortest (Table 2).

Participants using both SMS and paper generally rated the process of writing the text to respond as easy (a median rating of 4 for both groups), and there was no significant difference in this rating between modes, $t(157)=.77$, $p=.44$. Participants responding via paper using the original or SMS-version RTS instructions did not significantly differ in terms of RTS score mean or distribution ($t(21)=1.55$, p=.135; AD=1.83, p=.111). Due to this, analyses of the twenty item version of the RTS proceeded by pooling both original and new instructions under the category of 'paper' responses. ANOVA revealed there was also no significant difference in reported difficulty of responding between participants completing the different length versions via SMS, $F(3)=.062$, $p=.98$.

Table 2. Multi-group factor analysis outcomes

Version	Model 1 (Configural invariance)	Model 2 (Metric invariance)	Model 3 (Scalar invariance)	Model 4 (General Factor invariance)
5	$\chi^2(340)=12.483$, $p<=.254$	-	-	-
10	$\chi^2(70)=113$, $p<.001$	$\chi^2(79)=128$, $p<.001$	$\chi^2(88)=138$, $p<.001$	$\chi^2(89)=140$, $p<.001$
		$\delta\chi^2(9)=14$, $p=.09$	$\delta\chi^2(18)=25$, $p=.125$	$\delta\chi^2(19)=27$, $p=.09$
15	$\chi^2(180)=425$, $p<.001$	$\chi^2(194)=432$, $p<.001$	$\chi^2(208)=443$, $p<.001$	$\chi^2(209)=44$, $p<.001$
		$\delta\chi^2(14)=7.073$, $p=.904$	$\delta\chi^2(28)=18$, $p=.929$	$\delta\chi^2(29)=19$, $p=.922$
20	$\chi^2(340)=1019$, $p<.001$	$\chi^2(359)=1034$, $p<.001$	$\chi^2(378)=1067$, $p<.001$	$\chi^2(379)=1069$, $p<.001$
		$\delta\chi^2(19)=15$, $p=.713$	$\delta\chi^2(38)=48$, $p=.125$	$\delta\chi^2(39)=50$, $p=.112$

Discussion

Overall, cross-mode translation had little impact on the psychometric properties of the RTS. Ten, fifteen, and twenty item versions all conformed to the single factor structure established during the creation of the RTS by Brinker and Dozios (2009), and were consistently invariant across modes. The shorter versions provided no benefit in terms of response rates or psychometric validity, suggesting no significant detriment to administering the longer, twenty item version via SMS.

Only the shortest variant was problematic, demonstrating higher response rates at the cost of psychometric stability. The lack of equivalence across modes may reflect a genuine difference, as rumination does change with age (Sütterlin, Paap, Babic, Kübler, & Vögele, 2012). It is most likely that the lack of equivalence between the five item versions of the RTS administered on paper and via SMS is an artefact of the combination of having a relatively small sample, and having so few items available for analyses. Smaller samples are generally less stable than larger ones, as they allow more scope for outliers to distort mean tendencies and distributions. Instruments with more items tend to be more stable than their counterparts for the same reason. As factor analysis is sensitive to these issues, factor models become more stable (and thus more useful) with an increasing sample size, and an increasing number of items loading on the proposed latent factors (Tabachnick & Fidell, 2007). This is not always a concern, however the presence of low commonalities alongside small sample sizes, and small number of items loading on factors, inflates error terms and thus undermines the validity of the factor models (MacCallum & Widaman, 1999).

Though promisingly high response rates were diminished by missing items, all length variants compared favourably to current estimates of approximately 40% response rates for online self-report research (Cook, Heath, & Thompson, 2000; Shih, 2008). Response rates were notably consistent across the longer versions, suggesting that the presumed increase in burden of a twenty, rather than ten or fifteen, item questionnaire is not sufficient to produce a

significant impact on response rates. Results supported adapting response instructions to an SMS format with minimal disruption to the psychometric properties of responses. The general form of instructions used was effective.

Study 2

This second study extends on the previous by exploring cross-modal measurement invariance between SMS and online instruments, rather than paper. Online surveys have become the dominant mode in self-report research (Dillman et al., 2009). With the ubiquity of computer ownership and internet access, online questionnaires are beneficial as they tend to attract a broader sample than just university undergraduates participating in return for research credit (Gosling, Vazire, Srivastava, & John, 2000). This may also be the case for SMS research. Whilst the volume of psychology research undertaken with undergraduate participants makes it a good start point for examine the properties of a research mode, findings based upon an undergraduate sample has limited generalizability to the general population (Sears, 1986), in particular due to undergraduates having generally poorer mental health (Stallman, 2010). This study will therefore, in part, evaluate how well SMS performs when used with the general population.

Study 1 remained within the bounds of the longest instruments previously administered via SMS. The strong cross-mode equivalence of even the full twenty item RTS suggests further scope for pushing the boundaries of length when administering an instrument via SMS. As longer instruments tend to have more complex underlying factor structures, this allows exploration of cross-mode equivalence in scales with multiple factors. This will provide more scope for cross-mode differences in factor structures, loadings, means and intercepts, and thus prove a more stringent evaluation of the measurement invariance of administering an instrument via SMS.

Participants and materials

A total of 183 participants (57 via SMS, 126 online) were recruited to complete the 10-item negative axis of the Positive Affect Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988). Aged 17-64 (*M*=22), 57% of this sample were female. All obtained research credit as part of a larger study. The SMS version of the PANAS was administered in one single text, as follows.

> *Instructions: Copy each word, then indicate to what extent you feel this way right now, that is, at the present moment. An example correct response = bipedal 5. Use the following scale: 1 = very slightly or not at all, 2 = a little, 3 = moderately, 4 = quite a bit, 5 = extremely. The ten words are: distressed * upset * hostile * irritable * scared * afraid *ashamed * guilty * nervous * jittery. Reply to this text with your response. Please check you have completed all TEN words before sending your response. Issues? Contact [researcher's email] Thanks!*

A different sample of ` (36 via SMS, 217 online) were recruited to complete the 16-item Acceptance and Action Questionnaire (AAQ; Hayes et al., 2004). Aged 17-70 (*M*=24), 75% of this sample were female. Despite possible problems with using an undergraduate sample for comparison, practical reasons resulted in the online sample being undergraduate students participating in return for course credit. The SMS portion of the sample were members of the general public, offered no incentive. The AAQ was administered across three texts (one instruction text, two texts for the scale proper).

> *Instructions: In 15 minutes you will be sent a list of statements. Please rate the truth of each statement as it applies to you. Use the following scale to make your choice: 1 = never true, 2 = very rarely true, 3 = Seldom true, 4 = Sometimes true, 5 =*

Frequently True, 6 = Almost Always True, 7 = Always true. To respond, reply to the final SMS with the letter and truth of each statement, i.e. A3 B1 C4 etc

In fifteen minutes time, participants received the questions in the following format:

A. I am able to take action on a problem even if I am uncertain what is the right thing to do. B. A person who is really 'together' should not struggle with things the way I do [etc]

A total of 84 participants (57 via SMS, 27 online) were administered the 42-item Depression Anxiety and Stress Scale (DASS; Lovibond & Lovibond, 1995). Aged 17-62 (*M*=21), 58% were female. The DASS was administered across six texts (one instruction text, five for the scale proper), using the same completion instructions as the RTS in study 1 (i.e. 1.5x2.6 etc). All participants were undergraduates receiving an incentive of course credit.

A total of 391 participants (124 via SMS, 267 online) were administered the 60 item Positive Affect Negative Affect Schedule – Expanded form (PANAS-X; David Watson and Clark, 1999). Aged 18-46 (*M*=21), 65% were female. The PANAS-X was administered across two texts (one for instructions, and one with the scale proper). Instructions were similar to those for the PANAS, but required fewer characters in the reply to compensate for the drastically increased instrument length.

PANAS-X INSTRUCTIONS - This scale consists of a number of words and phrases that describe different feelings and emotions. Read each item and then mark the appropriate answer next to that word. Indicate to what extent you have felt this way during the past few weeks . Use the following scale to record your answers: 1= very slightly, 2 = a little, 3 = moderately, 4 = quite a bit, 5 = extremely. Reply with the

first two letters of each word, and your answer, i.e. for cheerful=5 and sad=2, ch5 sa2. The words/phrases will be sent to you in 15 mins.

To standardise the response experience as much as possible, online participants entered their responses for each item in text boxes, rather than select answers using a Likert matrix.

Design and procedure

This was a between-subjects correlational design. As in Study 1, those responding via SMS indicated their consent by texting "yes" and a time and date of their convenience to a number specified by the researcher, and were sent the instruction text at that time. The full questionnaires were then sent fifteen minutes later. Demographic information was collected in a follow-up text the next day. Those participating online indicated their consent via survey completion, and provided demographic information at the same time as completing the instrument.

Results

The response rate for the PANAS-NA was very high, as was the reliability of the scale completed via both modes. Responses completed via SMS had a significantly lower mean, and the distributions of the two modes significantly differed (table 3). This is due to a floor effect in SMS completion across most items – most SMS respondents answered with the lowest possible score for each item. As in the PANAS-NA norms (Watson et al., 1988), a single factor solution was sufficient for both modes. Multi-group factor analysis indicated configural but not metric invariance. Inspection of communalities reveals this is likely due to some problematic items in the SMS but not online factor structures. It is possible that the apparent high reliability and single factor structure from the SMS scores is due to the floor effect of consistently low scores across items, rather than reflecting the intended PANAS-NA

structure encapsulated in the more varied online responses. There was not a significant difference in age between those completing the PANAS-NA via SMS or online, $t(56)=1.94$, $p=0.06$.

Turning to the AAQ, reliability was similar, and there were no significant differences in mean and distributions of total scores by mode (table 3). However, there was a significant (over 50%) non-response rate in those responding via SMS. Subsequent factor analyses should be interpreted with caution due to the small SMS sample size. Having established that a single factor solution was sufficient, multi-group factor analysis revealed that while SMS and online completions of the AAQ had configural and metric invariance, they did not have scalar invariance (table 4).

Table 3. Missingness, descriptive and internal consistency comparisons between modes

	N	RR	UR	SMS			Online			t *p*	K
Scale				M	SD	α	M	SD	α		
PANAS-NA	10	99%	88%	13	4	.80	21	8	.88	10.33, *p*<.01	<.01
AAQ	16	46%	43%	68	9	.70	66	10	.71	1.15, *p*=.25	.52
DASS	42	96%	42%	44	24	.96					
(depression)				15	10	.93	8	7.6	.88	2.45, *p*=.02	<.01
(anxiety)				13	9	.90	7	7.7	.92	2.57, *p*=.01	.05
(stress)				17	8	.87	6	3	.91	5.081, *p*<.01	.02
PANAS-X	60	83%	57%								
Negative affect	10			23	8	.86	15	5	.85	9.6, p<.001	<.01
Positive affect	10			31	7	.83	27	7	.86	5.87, p<.001	<.01
Basic negative emotion scales											
(fear)	6			13	5	.85	8	3	.82	8.12, p<.001	<.01
(hostility)	6			12	4	.77	8	3	.78	8.92, p<.001	<.01
(guilt)	6			14	5	.86	10	4	.80	6.59, p<.001	<.01
(sadness)	5			12	5	.87	9	4	.83	5.22, p<.001	<.01
Basic positive emotion scales											
(joviality)	8			26	6	.87	21	7	.93	6.97, p<.001	<.01
(self-assurance)	6			16	4	.74	14	5	.85	5.41, p<.001	<.01
(attentiveness)	4			13	3	.65	12	3	.68	2.13, p=.033	.09
Other affective states											
(shyness)	4			8	3	.58	7	3	.73	4.233, p<.001	<.01

(fatigue)	4	14	4	.80	10	4	.87	8.36, p<.001	<.01
(serenity)	3	9	3	.84	10	3	.81	2.44, p=0.01	.03
(surprise)	3	6	2	.55	4	2	.72	8.51, p<0.01	<.01

Note. RR = response rate, the percentage of participants who at least partially completed the SMS instrument. UR = usable response rate, the percentage of participants who correctly completed all items of the SMS instrument. A *p*K value less than 0.05 signifies a significant result for the Anderson-Darling k-sample test, indicating the scale's total scores come from different underlying distributions depending on mode.

Table 4. Multi-group factor analysis outcomes

Scale	N items	N factors	Model 1 (Configural invariance)	Model 2 (Metric invariance)	Model 3 (Scalar invariance)	Model 4 (General F. invariance)
PANAS-NA	10	1	$\chi^2(70)=221$, $p<.001$	$\chi^2(79)=257$, $p<.001$	-	-
				$\delta\chi^2(9)=36$, $p=.001$	-	-
AAQ	16	1	$\chi^2(208)=507$, $p<.001$	$\chi^2(223)=530$, $p<.001$	$\chi^2(238)=583$, $p<.001$	-
				$\delta\chi^2(15)=22$, $p=.1$	$\delta\chi^2(30)=76$, $p=.001$	-
DASS	42	3	$\chi^2(1632)=9939$, $p<.001$	$\chi^2(1671)=10054$, $p<.001$	-	-
				$\delta\chi^2(39)=115$, $p<.001$	-	-
PANAS-X (general dimensions)	60 (20)	2	$\chi^2(338)=847$, $p<.001$	$\chi^2(356)=864$, $p<.001$	$\chi^2(374)=990$, $p<.001$	-
				$\delta\chi^2(18)=17$, $p=.501$	$\delta\chi^2(36)=143$, $p=.001$	-
PANAS-X (subscales)	60 (40)	11	$\chi^2(2642)=5548$, $p<.001$	$\chi^2(2686)=5626$, $p<.001$	-	-
				$\delta\chi^2(44)=77$, $p=.001$	-	-

Across all three subscales, online completions of the DASS resulted in consistently, significantly lower scores, coupled with a more skewed, zero-bounded distribution than SMS completions (table 2). Online completions were closer to norms reported for a general population, whilst SMS were descriptively closer to norms reported for clinical populations (Lovibond & Lovibond, 1995). There was no significant difference in the age of those participating via online survey, or SMS $t(27)=1.87$, $p=.07$. As in the literature, a three factor solution (corresponding to depression, stress and anxiety subscales) was sufficient. Multi-group factor analysis revealed configural, but not metric invariance.

Despite an initially high response rate for the PANAS-X, widespread item missingness in SMS responses considerably diminished the number of usable responses. Despite this missingness, reliability was again relatively high across both SMS and online administrations. Showing an opposite pattern to the PANAS-NA, mean scores for SMS completions were higher than online completions for all but one PANAS-X subscale (the serenity subscale) (Table 2). The online PANAS-X scores were consistently closer to the norms reported by Watson and Clark (1999) than the SMS scores. There was a significant difference in age between groups, $t=5.3091$, $p<.01$, with SMS respondents being younger (*M* age 20) than online respondents (*M* age 23).

Due to the large number of factors present in the original scale, and floor effects in the online conditions, multi-group factor analysis of all sub-scales failed to converge. Analysis proceeded by dividing investigation into the 'classic' 20 item PANAS-NA/PA two factor structure, and the additional sub-scales established in the PANAS-X (as in Watson & Clark, 1999). The expected two-factor structure was appropriate for PANAS-NA and PANAS-PA items, with multi-group factor analysis indicating configural, metric, but not scalar invariance. Whilst the proposed factor structure for the other sub-scales was theoretically feasible, the model was too unstable for conclusions to be drawn.

Discussion

Overall, cross-mode translation significantly impacted upon the psychometric properties of the instruments in terms of means and latent structure. The PANAS-NA proved problematic in terms of cross-mode equivalence, with SMS responses considerably differing from their online counterparts, and what would be expected given the PANAS-NA norms (Watson et al., 1988). Though age may have played a factor in the cross-mode differences in the five item RTS in study 1, age of participants completing the PANAS-NA did not significantly differ by mode. Nor could the different scores be explained by differential social desirability effects stemming from contact with the researcher, as participants using both modes had the same contact with the researcher. Social desirability may have impacted responses as a function of response context. Due to their portability mobile phones are commonly used in public places (Wei & Leung, 1999), whilst online questionnaires are commonly attempted on more stationary personal computers in research laboratories, workplaces, or the home. This means that a participant using SMS for the purposes of research is more likely to receive the instrument when in a public setting (such as in a shopping mall, or in a restaurant) than one responding online. Perhaps participants were unwilling to honestly disclose the degree of their negative affective state whilst in a public setting, hence in the current study under-reported their negative affect. It is possible this effect is absent from the longer SMS instruments, because participants were willing attempt the briefer measure as the SMS arrived, but removed themselves from the company of others to focus on responding to the longer instruments. Future research could investigate this possibility by adding questions about current participant location and social surroundings at to the SMS instrument.

Just as past cross-mode investigations found web questionnaires tended to have systematically higher scores in general (Vecchione et al., 2012), so SMS scores were higher for the AAQ, DASS, and PANAS-X. These higher scores via SMS pulled the average away

from zero, resulting in scores conforming to a more normal distribution shape than paper scores. Unexpectedly, the 16 item AAQ was the only scale demonstrating measurement invariance in terms of response mean and distributions. This is bemusing given that the comparison was between the general population and undergraduate students, and there is reason to believe the undergraduate samples should obtain poorer scores on instruments relating to mental health (Sears, 1986; Stallman, 2010). This apparent equivalence may be a consequence of the small SMS comparison group. In a review of response rates to paper questionnaires, Heberlein and Baumgartner (1978) noted that surveys administered to the general population are less likely to be returned than those given to specific subsamples, such as student populations. A meta-analysis of online survey response rates found that using an academic (including students and faculty) sample constituted a suppressor variable for other factors impacting on response rates, such as offering monetary incentives (Cook et al., 2000). Perhaps the comparably high response rate to SMS instruments administered to undergraduate samples is specific to that group, especially where course credit (an incentive unique to university student participants) is offered. Strategies known to improve response rates, such as monetary or token gift incentives (Dillman et al., 2009), may be required when administering instruments via SMS in the general population.

Cross-modal equivalence was only partially demonstrated in the instruments with multiple underlying factors. Support for metric and configural but not scalar or general invariance, suggests that though the general factor structure is preserved across modes, underlying distribution and mean-level differences translate into non-equivalent latent mean scores and intercepts. This is similar to previous findings that online administrations tend to have higher latent mean scores than paper administrations (Cole, 2006; Meade et al., 2007). It is not likely due to age differences between SMS and online samples; there was no difference in age for DASS completions, and though SMS PANAS-X respondents were statistically

significantly younger than online respondents, the practical difference was small (just three years difference in mean age).

General discussion

The extent to which administration via SMS impacted upon the psychometric properties of pre-existing psychological instruments was somewhat associated with the length of the instrument. In study 1, cross-mode translation had little impact on the psychometric properties of the RTS, except for the shortest length variant, supporting the validity of administering the full twenty-item instrument via SMS. Study 2 revealed a similar pattern detected by research comparing paper and online administrations (Cole, 2006; Meade et al., 2007; Vecchione et al., 2012): though the factor structures were congruent with what would be expected from instrument norms, SMS scores were higher than their online counterparts for the AAQ, DASS, and PANAS-X. Compounded by distribution differences, this ultimately led to non-equivalence of the underlying factor means and intercepts. These differences were not due to age, and thus differential access to and aptitude with SMS technology. This, and the surprising mean-level equivalence of the instrument administered using different modes to different samples, suggests that the cross-mode variance present in administration via SMS stems from a property of the response mode itself, rather than different populations engaging with the instruments in different ways. If participants were responding via SMS in more public contexts than those responding via online or paper, and this caused under-reporting of negative attributes due to social desirability bias, one would expect SMS scores to be lower than those from other modes.

Another possible explanation relates to the visual layout of the instruments. When participants respond using a likert matrix, they are presented with a vertical list of questions and a horizontal array of tickable boxes corresponding to the response. In this way, the full

range of possible responses for each question is always visible. In the first study, cross-mode equivalence was found when both SMS and paper respondents were asked to respond by writing their answers in a string of text, with neither mode using a likert matrix. In the second study, participants completing the AAQ and DASS online responded via a likert matrix, whilst those responding via SMS responded in text strings. The SMS responses therefore had no spatial representation of the range upon which to anchor their responses. This lack of an implicit visual anchor may have led to systematic differences in responses. However, despite this difference the AAQ scores were equivalent. Further, to control for such effects, the online administration of the PANAS-X did not use a likert matrix for responses, yet still demonstrated consistently lower scores than the SMS administration. It is therefore unlikely that this visual layout cue underlies the difference in scores found across modes.

In terms of response rates; different target populations engaged with SMS differently. In instruments administered to undergraduate samples, response rates to even the longer instruments exceeded expectations, but the response rate from the scale administered to the general population was on par with the 40% response rate estimated for online self-report research (Cook, Heath, & Thompson, 2000; Shih, 2008). Across all samples, missingness within responses was far more of a concern than response rates, considerably diminishing the amount of usable data. Degree of missingness was not linearly related to instrument length, as there was a sudden jump from low to high missingness between the shortest scale, and the others in each study. This is not likely due to possible environmental distraction when completing an SMS measure, as that would manifest in missingness scaling with the number of items (as each additional item provides another opportunity for a distraction to cause a missed response). The more probable explanation is the number of SMS needed to administer the questionnaire. Both the five item version of the RTS, and the PANAS-NA, could be sent in one single text. Responding therefore required less navigation between SMS, lowering

participant burden and making it easier to read and remember the questions, thus improving response behaviour. This possibility needs to be explored in more depth, however, as participants responding to the longer instruments in the first study did not report the response process as any more difficult. Indeed, participants in study one generally reported the process of typing their responses easy, regardless of the length of the questionnaire. In study 2, there was no evidence of mode impacting on internal reliability, as would have been the case if the relatively inaccurate nature of typing on a smartphone rather than computer keyboard (Page, 2013) had begat typographic errors.

These results demonstrate that an instrument as long as sixty items in length can be administered by SMS, with comparable response rates, internal reliability, and factor structures to online administration. However, in instruments over ten items in length, mean responses tended to be higher, leading to lack of equivalence in terms of latent means and intercepts. Measurement invariance was strongest in the instruments with a single underlying factor. There is much that could be done to expand upon this preliminary investigation. The difference between instrument lengths used here were relatively coarse; the issues discussed here could be examined in much greater detail by systematically, repeatedly adding a single item to a questionnaire and examining cross-mode equivalence at each step. There also remains the question as to why SMS scores tend to be considerably higher than their online counterparts. These results tentatively suggest it is not age, instrument length (beyond ten items), distraction, or difficulty typing out the response – if this is supported by future research, additional explanations will be required.

References

Ainsworth, J., Palmier-Claus, J. E., Machin, M., Barrowclough, C., Dunn, G., Rogers, A., Buchan, I., et al. (2013). A comparison of two delivery modalities of a mobile phone-based assessment for serious mental illness: native smartphone application vs text-messaging only implementations. *Journal of medical Internet research*, *15*(4), e60. doi:10.2196/jmir.2328

Banbury, S., & Macken, W. (2001). Auditory distraction and short-term memory: Phenomena and practical implications. *Human Factors*, *43*(1), 12 – 29. doi: 10.1518/001872001775992462

Bexelius, C., Merk, H., Sandin, S., Ekman, A., Nyrén, O., Kühlmann-Berenzon, S., Linde, A., et al. (2009). SMS versus telephone interviews for epidemiological data collection: feasibility study estimating influenza vaccination coverage in the Swedish population. *European journal of epidemiology*, *24*(2), 73–81. doi:10.1007/s10654-008-9306-7

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual review of psychology*, *54*(1), 579–616. doi:10.1146/annurev.psych.54.101601.145030

Borsboom, D. (2006). When does measurement invariance matter? *Medical care*, *44*(11 Suppl 3), S176–81. doi:10.1097/01.mlr.0000245143.08679.cc

Brinker, J. K., & Dozios, D. J. A. (2009). Ruminative Thought Style and Depressed Mood. *Journal of Clinical Psychology*, *65*(1), 1–19. doi:10.1002/jclp.20542

Burchell, B., & Marsh, C. (1992). The effect of questionnaire length on survey response. *Quality and Quantity*, *26*(3), 233–244. doi:10.1007/BF00172427

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

Cole, M. S. (2006). The Measurement Equivalence of Web-Based and Paper-and-Pencil Measures of Transformational Leadership: A Multinational Test. *Organizational Research Methods*, *9*(3), 339–368. doi:10.1177/1094428106287434

Cook, C., Heath, F., & Thompson, R. L. (2000). A Meta-Analysis of Response Rates in Web- or Internet-Based Surveys. *Educational and Psychological Measurement*, *60*(6), 821–836. doi:10.1177/00131640021970934

Cook, T. D., & Campbell, D. T. (1979). *Quasi-Experimentation Design and Analysis issues for Field Settings* (pp. 37 – 95). Boston: Houghton Mifflin Company.

De Beuckelaer, A., & Lievens, F. (2009). Measurement Equivalence of Paper-and-Pencil and Internet Organisational Surveys: A Large Scale Examination in 16 Countries. *Applied Psychology*, *58*(2), 336–361. doi:10.1111/j.1464-0597.2008.00350.x

De Lepper, A. M., Eijkemans, M. J. C., Van Beijma, H., Loggers, J. W., Tuijn, C. J., & Oskam, L. (2013). Response patterns to interactive SMS health education quizzes at two sites in Uganda: a cohort study. *Tropical medicine & international health*, *18*(4), 516–21. doi:10.1111/tmi.12059

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

Fuchs, M. (2009). Differences in the Visual Design Language of Paper-and-Pencil Surveys Versus Web Surveys: A Field Experimental Study on the Length of Response Fields in Open-Ended Frequency Questions. *Social Science Computer Review*, *27*(2), 213–227. doi:10.1177/0894439308325201

Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2000). Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. *The American psychologist*, *59*(2), 93–104. doi:10.1037/0003-066X.59.2.93

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.

Hayes, S. C., Strosahl, K., Wilson, K. G., Bisset, R. T., Pistorello, J., Toarmino, D., Polunsy, M. A., et al. (2004). Measuring Experiential Avoidance: A Preliminary Test of a Working Model. *The Psychological Record*, *54*(4), 553 – 578. doi: 10.1002/jclp.20400

Heberlein, T. A., & Baumgartner, R. (1978). Factors affecting response rates to mailed questionnaires: a quantitative analysis of the published literature. *American sociological review*, *43*(4), 447 – 462. doi: 10.2307/2094771

Herrero, J., & Meneses, J. (2006). Short Web-based versions of the perceived stress (PSS) and Center for Epidemiological Studies-Depression (CESD) Scales: a comparison to pencil and paper responses among Internet users. *Computers in Human Behavior*, *22*(5), 830–846. doi:10.1016/j.chb.2004.03.007

Hirai, M., Vernon, L. L., Clum, G. a., & Skidmore, S. T. (2011). Psychometric Properties and Administration Measurement Invariance of Social Phobia Symptom Measures: Paper-Pencil vs. Internet Administrations. *Journal of Psychopathology and Behavioral Assessment*, *33*(4), 470–479. doi:10.1007/s10862-011-9257-2

Jansen, J. (1985). Effect of questionnaire layout and size and issue-involvement on response rates in mail surveys. *Perceptual and Motor Skills*, 139–142. doi:10.2466/pms.1985.61.1.139

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

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

Lagerros, Y. T., Sandin, S., Bexelius, C., Litton, J.-E., & Löf, M. (2012). Estimating physical activity using a cell phone questionnaire sent by means of short message service (SMS): a randomized population-based study. *European journal of epidemiology*, *27*(7), 561–6. doi:10.1007/s10654-012-9708-4

Lee, S. S. S., Xin, X., Lee, W. P., Sim, E. J., Tan, B., Bien, M. P. G., Lau, A. S. T., et al. (2013). The feasibility of using SMS as a health survey tool: an exploratory study in patients with rheumatoid arthritis. *International journal of medical informatics*, *82*(5), 427–34. doi:10.1016/j.ijmedinf.2012.12.003

Leung, D. Y. P., & Kember, D. (2005). Comparability of Data Gathered from Evaluation Questionnaires on Paper and Through the Internet. *Research in Higher Education*, *46*(5), 571–591. doi:10.1007/s11162-005-3365-3

Lim, M. S. C., Sacks-Davis, R., Aitken, C. K., Hocking, J. S., & Hellard, M. E. (2010). Randomised Controlled Trial of Paper, Online and SMS Diaries for Collecting Sexual Behavior Information from Young People. *Journal of Epidemiology and Community Health*, *64*(10), 885–889. doi:10.1136/jech.2008.085316

Lovibond, S. H., & Lovibond, P. f. (1995). *Manual for the depression anxiety stress scales* (2nd ed., Vol. 65). Sydney: Psychology foundation.

MacCallum, R., & Widaman, K. (1999). Sample size in factor analysis. *Psychological methods*, *4*(1), 84 –99. 10.1037/1082-989X.4.1.84

Mackay, M. M., & Weidlich, O. (2009). Austrailan Mobile Phone lifestyle index. *Specialist*. Australian Interactive Media Industry Association Mobile Industry Group.
Meade, A. W., Michels, L. C., & Lautenschlager, G. J. (2007). Are Internet and Paper-and-Pencil Personality Tests Truly Comparable?: An Experimental Design Measurement Invariance Study. *Organizational Research Methods*, *10*(2), 322–345. doi:10.1177/1094428106289393
Page, T. (2013). Usability of text input interfaces in smartphones. *Journal of design research*, *11*(1), 39–56. 10.1504/JDR.2013.054065
Richardson, C., & Johnson, J. (2009). The influence of web- versus paper-based formats on the assessment of tobacco dependence: evaluating the measurement invariance of the Dimensions of Tobacco Dependence Scale. *Substance abuse: research and treatment*, *3*, 1–14. doi: 10.1111/j.1365-2702.2009
Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review*, *18*(4), 210–222. doi:10.1016/j.hrmr.2008.03.003
Sears, D. (1986). College sophomores in the laboratory: Influences of a narrow data base on social psychology's view of human nature. *Journal of personality and social psychology*, 515–530. doi: 10.1037/0022-3514.51.3.515
Shapiro, J. R., Bauer, S., Hamer, R. M., Kordy, H., Ward, D., & Bulik, C. M. (2008). Use of text messaging for monitoring sugar-sweetened beverages, physical activity, and screen time in children: a pilot study. *Journal of nutrition education and behavior*, *40*(6), 385–91. doi:10.1016/j.jneb.2007.09.014
Shih, T.-H. (2008). Comparing Response Rates from Web and Mail Surveys: A Meta-Analysis. *Field Methods*, *20*(3), 249–271. doi:10.1177/1525822X08317085
Smith, T. W. (1993). Little things matter: A sampler of how differences in questionnaire format can affect survey responses. *GSS Methodological report*.
Stallman, H. M. (2010). Psychological distress in university students: A comparison with general population data. *Australian Psychologist*, *45*(4), 249–257. doi:10.1080/00050067.2010.482109
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
Sütterlin, S., Paap, M. C. S., Babic, S., Kübler, A., & Vögele, C. (2012). Rumination and age: some things get better. *Journal of aging research*, *2012*, 267327. doi:10.1155/2012/267327
Tabachnick, B. G., & Fidell, L. S. L. S. (2007). *Using Multivariate Statistics* (Fifth.). USA: Pearson.
Vandenberg, R. J., & Lance, C. E. (2000). A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research. *Organizational Research Methods*, *3*(1), 4–70. doi:10.1177/109442810031002
Vecchione, M., Alessandri, G., & Barbaranelli, C. (2012). Paper-and-pencil and web-based testing: the measurement invariance of the Big Five personality tests in applied settings. *Assessment*, *19*(2), 243–6. doi:10.1177/1073191111419091
Watson, D, Clark, L. A., & Tellegen, a. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology*, *54*(6), 1063–70. doi: 10.1037/0022-3514.54.6.1063
Watson, David, & Clark, L. (1999). The PANAS-X: Manual for the positive and negative affect schedule-expanded form. doi: 10.13072/midss.438
Wei, R., & Leung, L. (1999). Blurring public and private behaviors in public space: policy challenges in the use and improper use of the cell phone. *Telematics and Informatics*, *16*(1-2), 11–26. doi:10.1016/S0736-5853(99)00016-7
Whitford, H. M., Donnan, P. T., Symon, A. G., Kellett, G., Monteith-hodge, E., Rauchhaus, P., & Wyatt, J. C. (2012). Evaluating the reliability , validity , acceptability , and practicality of SMS text messaging as a tool to collect research data : results from the Feeding Your Baby project. *J Am Med Inform Assoc*. doi:10.1136/amiajnl-2011-000785
Wicherts, J. M., Dolan, C. V, & Hessen, D. J. (2005). Stereotype threat and group differences in test performance: a question of measurement invariance. *Journal of personality and social psychology*, *89*(5), 696–716. doi:10.1037/0022-3514.89.5.696
Yu, S.C., & Yu, M.-N. (2007). Comparison of Internet-based and paper-based questionnaires in Taiwan using multisample invariance approach. *Cyberpsychology & behavior : the impact of the Internet, multimedia and virtual reality on behavior and society*, *10*(4), 501–7. doi:10.1089/cpb.2007.9998