

Survey Development for Adolescents Aged 11 to 16 Years: A Developmental Science Based Guide

Key words: Survey development, adolescents, satisficing, respondent characteristic, question characteristic

Authors: Atefeh Omrani¹, Joanna Wakefield-Scurr², Jenny Smith³ & Nicola Brown¹

1. School of Sport Health and Applied Science, St Mary's University, Waldegrave Road, Twickenham, TW1 4SX, UK
2. Department of Sport and Exercise Science, University of Portsmouth, Spinnaker Building, Cambridge Road, Portsmouth, PO1 2ER, UK
3. Department of Sport and Exercise Sciences, University of Chichester, Bishop Otter Campus, College Lane, Chichester, West Sussex, PO19 6PE, UK

Corresponding author: Atefeh Omrani, School of Sport, Health & Applied Science, St Mary's University, Waldegrave Rd, Twickenham, TW1 4SX, UK

T: 02082408244

F: 0044 2082 404212

E: 157483@live.stmarys.ac.uk

Abstract

Methodological knowledge on surveying young adolescents is scarce and researchers often rely on theories and methodological studies based on adult respondents. However, young adolescents are in the process of developing their cognitive, psychological, emotional and social skills, therefore present a unique set of considerations. Question characteristics including; question type and format, question difficulty, wording, ambiguity, the number of response options, and the inclusion of a neutral mid-point, play a pivotal role in the response quality of young adolescents. Failure to address these factors is likely to encourage young adolescents to use satisficing techniques. In this article, we provide a science based guide for developing surveys for use with adolescents aged 11 to 16 years. The guide considers the characteristics and developmental stages of adolescents as survey responders and incorporates advice on appropriate question characteristics, survey layout and question sequence, approaches to pre-testing surveys and mode of survey administration. The guide provides recommendations for developmentally appropriate survey design to improve response quality in survey research with young adolescents.

Introduction

Social science researchers increasingly survey young adolescents (Borgers et al. 2000; Borgers et al. 2004; Scott 1997). However, methodological knowledge on how to survey young adolescents is scarce, and much advice is derived from methodological studies on adults and theories based on adult respondents (Arthur et al. 2017; Borgers et al. 2000; Borgers et al. 2004). Procedures to enhance response quality and improve data collection methods in adult populations are well documented (Borgers et al. 2000; Bryman 2012; De Vaus 2014; Moser and Kalton 1985; Robson 2011). However, they may not be appropriate for young adolescents due to differences in characteristics of this population and adult populations (Borgers et al. 2000; Borgers et al. 2003; Borgers et al. 2004; de Leeuw 2011; Vaillancourt 1973). As a transitional period between childhood and adulthood, adolescence involves rapid developmental changes in physical, cognitive, social, and emotional development (Leshem 2016). There are significant developmental differences between adolescents and other age groups, which need to be taken into account in survey research with adolescents (Borgers et al. 2000). Compared to children and adults, adolescents tend to be good at seeking out new experiences, meeting new people, recognizing social and emotional information, and confronting various challenges (Leshem 2016). Cognitive abilities increase with age during adolescence, which accelerates the acquisition of knowledge and learning of cognitive and emotional skills (Reyna and Rivers 2008). However, normative development of adolescents is also characterized by sensitivity to reward, emotionality, increased risk-taking, lower self-management and impulsivity, with a tendency to act in the spur of the moment and to make rash choices regardless of the consequences (Leshem 2016; Scott and Steinberg 2008). These psychosocial and emotional factors contribute to immature judgment and decision making skills in adolescence (Scott and Steinberg 2008). Logical decision making ability increases with age, stabilizing around age 21 (Cauuffman and Steinberg 2000; Partridge 2010). The differences in adolescents and adult characteristics described above have profound implications for survey research with adolescents and

should be carefully considered in developing surveys for use with this population. Question wording, structure and length of surveys are important factors for data quality as is the developmental phase of adolescents (de Leeuw 2011).

There are further developmental differences between adolescents and other age groups. During middle childhood the structure of self-concept changes, and in late middle childhood (10 to 12 years) children start comparing themselves with others (de Leeuw 2011). During early adolescence peer relationships take on a heightened importance, increasing the likelihood of adolescents changing their decisions or behavior in response to group norms and peer pressure (de Leeuw 2011; Leshem 2016; Somerville 2013). When developing surveys for adolescents, extreme care should be paid to group norms and peer pressure to obtain better data quality (de Leeuw 2011).

An increasing body of empirical evidence suggests that respondent and question characteristics affect the reliability of survey responses (Borgers et al. 2000; Borgers and Hox 2000; Borgers et al. 2003; 2004; Fuchs 2005; Greig et al. 2013; Menold et al. 2014). Krosnick's satisficing theory (1991) explains the interaction effect between respondent and question characteristics, which offers an explanation for differences in reliability of responses between respondents, and for the effects of question characteristics. The satisficing theory was developed based on a standard question answering process-model, originally developed by Tourangeau (Tourangeau et al. 2000). Tourangeau believed that in order to optimally answer survey questions, respondents should move through four stages of cognitive processing: comprehension, recall, judgement and response. Satisficing means that a respondent gives superficial responses that appear reasonable or acceptable, without going through all the stages involved in the question answering process. According to the satisficing theory, there is a strong relationship between the respondents' cognitive ability and the reliability of responses. It assumes that less cognitively sophisticated respondents are more sensitive to cognitively demanding questions, and this can have a negative effect on the reliability of responses (Krosnick 1991; Krosnick and Presser 2010). General

determinants of satisficing include the motivation of the respondent, the difficulties of the task, and the cognitive abilities of the respondent (Borgers et al. 2003; Krosnick 1991). Respondents with lower cognitive ability simply provide a satisficing answer (e.g. responding don't know, giving superficial responses or randomly choosing a response), when optimally answering a survey question requires substantial cognitive effort (Krosnick 1991; Krosnick and Presser 2010). Applying the satisficing theory to children or adolescent populations is of special interest of survey researchers, because growing up and aging involve changes in cognitive functioning, and cognitive ability is a central respondent characteristic that affects the reliability of responses (Borgers et al. 2003; 2004).

Designing a suitable survey that targets young adolescents requires a balance between their characteristics (such as age), developmental stages and question characteristics (Alwin 2007; Borgers et al. 2003; 2004; Krosnick 1991). Data quality in survey research with adolescents depends highly on the age and cognitive development of the adolescents (Mavletova and Lynn, 2017). Gender and age are often used as proxy variables for cognitive development (Borgers et al. 2000). Borgers et al. (2000) found that both gender and age had an influence on the reliability of responses in survey research with children and adolescents. Young children (7 to 10 years), who are beginning their education, produce less reliable and consistent responses on surveys. Children of this age tend to please the researcher or teacher, which may result in more superficial answers and socially desirable responses (de Leeuw 2011). At ages 11 to 12 years, the reliability and consistency of responses increases, stabilizing around 14 to 15 years. Adolescents aged 11 and older have most skills needed to answer nearly every type of question and complete surveys with minimal errors related to developmental concerns (Arthur et al. 2017). In general, girls produce slightly more reliable responses and better quality data than boys, irrespective of age (Borgers et al. 2000; Mavletova and Lynn 2017).

The developmental capacity to respond to questions impacts response quality. Therefore, the stages of cognitive development should be considered when developing a survey for adolescents (Borgers et al. 2000). The most widely-known and accepted theory of cognitive development is Piaget's stage theory, which suggests that children move through four different stages of mental development, Sensorimotor (birth to age 2), pre-operational (2 to 7 years), concrete operational (7 to 11 years), and formal operational (≥ 11 years) (Feldman 2004; Piaget 1929). The timing of the stages have received some criticism as they tend to overlap and the transitions from one stage to another are not as clear as assumed (Borgers et al. 2000; Borgers et al. 2004). However, Piaget's (1929) theory can provide survey researchers with information about the basic levels and directions of cognitive development. In combination with the question answering model, and the satisficing theory, it can give researchers an understanding of how children and adolescents differ from adults, and explain why they have more difficulties with cognitive demanding survey questions than adults (Borgers et al. 2004; Scott 1997). The developmental stage of young adolescents aged 11 to 16 is labelled as formal operational (Piaget 1929; Turner 1975). At the formal operational stage, cognitive functioning and social skills are quite well developed and abstract thought emerges. However, adolescents at this age are very context sensitive, literal, and liable to lose their motivation and get easily bored (Borgers et al. 2000; Scott 1997). Failure to consider these factors when developing survey questions for adolescents is likely to increase the likelihood of young adolescents using satisficing techniques and not go through all the stages of the question answering model (Borgers et al. 2000; Borgers et al. 2004; Krosnick 1991). Consequently, question characteristics play a pivotal role in the response quality (de Leeuw et al. 2004; Greig et al. 2013; Smith and Platt 2013; Vaillancourt 1973). Question characteristics that should be carefully considered when designing survey questions for adolescents include question type and format, question difficulty, wording, ambiguity, the number of response options, and whether to include a neutral mid-point on rating scales.

The aim of this article is to summarize a wide range of literature in the area of survey development, and apply the theoretical background discussed above, to present an inclusive survey development guide for use with adolescents aged 11 to 16 years. The guide can be applied to various survey measures including knowledge, attitudes, emotion, cognition, intention and behavior. The intention is to provide information to researchers who may undertake research requiring such an approach. The guide incorporates important considerations including selecting suitable question types, determining wording and length of questions, selecting appropriate question formats, deciding on appropriate number of points on rating scales, labelling of rating scale response categories, and the use of negative or positive wording. Additionally, the guide provides advice on survey layout and question sequence, approaches to pre-testing surveys and mode of survey administration. Key considerations for survey development in young adolescents are summarized in Table One. It must be acknowledged that although key considerations for survey development for use with adolescents aged 11 to 16 years are included within the article, more topics (e.g. analyzing, interpreting survey responses, behavior coding and order of response categories) could have been covered.

Type of questions (open versus closed questions)

Type and format of questions in a survey affect the quality of data collected, therefore this needs to be considered in the early stages of survey development (DeVellis 2016; Krosnick and Presser 2010). Firstly, when designing a survey a decision needs to be made whether to use open or closed questions. Close-ended questions limit the respondent to a set of fixed responses (Rea and Parker 2005; Reja et al. 2003). Conversely, open questions allow the respondent to express their opinion freely, adding richness to survey results and often providing more reliable and valid measurement than closed items (Krosnick and Presser 2010). Nevertheless, the satisfactory completion of open-ended questions requires a greater degree of cognitive ability and communicative skills compared to close-ended questions, impacting the reliability of responses (Bradburn et al. 2004; Rea and Parker 2005; Rossi et

al. 2013; Seale 2004). In addition, the probability of 'don't know' answers, and item non-response, is higher in open-ended questions compared to closed questions (Dillman et al. 2014; Krosnick and Presser 2010; Ogan et al. 2013; Reja et al. 2003; Rossi et al. 2013). This can be even more problematic with young adolescents as the increased difficulty of completing open-ended questions can easily encourage this population to apply satisficing techniques (Borgers and Hox 2000; de Leeuw 2011). Moreover, analyzing and coding answers to open-ended questions can involve an extensive amount of time and money, increasing the burden on researchers (Rea and Parker 2005; Reja et al. 2003).

In adult populations, the practical disadvantages of open-ended questions stated above, do not apply to the measurement of numeric quantities, for example money spent for goods or number of doctor visits (Krosnick and Presser 2010). However, this is not the case when surveying young adolescents, due to the difficulties they face retrieving information from memory without any prompts available (Borgers et al. 2000; Smith and Platt 2013). Smith and Platt (2013) employed an experimental design to compare close-ended and open-ended response categories for questions on frequency and quantities with a large-scale sample of adolescents aged 11-16 (n = 2,739). They concluded that better quality data is provided by close-ended questions rather than open-ended response options. Research has identified that this is also the case when measuring categorical judgement (such as the 'most important problem'), and testing factual knowledge (Bradburn et al. 2004; Rea and Parker 2005). Therefore, applying close-ended questions is preferable to the use of open-ended questions for young adolescents. This will minimize task difficulty and maximize adolescents' motivation to participate, thus reducing the likelihood of applying satisficing techniques (Arthur et al. 2017; De Vaus 2014; Krosnick and Presser 2010; Smith and Platt 2013). Open-ended questions should be used sparingly and only when absolutely necessary (Rea and Parker 2005).

Due to the self-administrating nature of surveys, probing to expand on any response is not possible; this may be particularly problematic for open-ended questions (Reja et al. 2003). When attempting to gather in-depth rich information from open ended questions, young adolescents may use satisficing techniques, or not respond due to the increased cognitive burden, especially because no prompts or probing are available (Smith and Platt 2013). If survey researchers require more qualitative information, utilizing surveys in combination with other methods in which probing or prompting is possible, such as focus groups and interviews, are recommended (Kitzinger 1995; Morgan 1996; Robson 2011).

Question wording and length

Adolescents are still developing their decisional and thinking skills. Therefore, to ease the question answering process in this population and produce better quality data, questions for young adolescents should be written in a clear and understandable way (e.g. using simple words and avoiding double barreled items) to avoid any ambiguity (de Leeuw 2011; Bell 2007; DeVellis 2016; Dillman et al. 2014; Kelley et al. 2003; Robson 2011). Survey questions should be kept as simple and as concrete as possible (de Leeuw 2011). They should also be short (less than 20 words in length), as long questions cause a lack of motivation and difficulties in concentration which will result in poorer data quality (Arthur et al. 2017; Borgers et al. 2000; Borgers and Hox 2000; Boynton and Greenhalgh 2004; De Vaus 2014). In addition, the overall length of a survey should be short, as long surveys increase the burden on participants which leads to increased reluctance to participate and higher non-response rates (De Vaus 2014; Rea and Parker 2005). In general, it has been shown that when survey length is short (≤ 10 to 20 minutes) response rates will be higher (Deutskens et al. 2004; Galesic and Bosnjak 2009; Marcus et al. 2007). When surveying young adolescents, it is recommended that the length of survey should be even shorter (for example, ≤ 10 minutes) to enhance response rates, and the willingness to complete the survey (Borgers and Hox 2000; 2001).

Question wording and length can affect the readability of a survey (Lenzner 2014) and it is recognized that survey respondents may not comprehend items with readability scores that exceed their reading ability (Calderón et al. 2006). Questions with a high readability index (easy to read) improve the reliability of responses in survey research with children and young adolescents (Borgers and Hox 2000; 2001). Therefore, a readability formula should be utilized to ensure that the survey questions are appropriate for the reading level of young adolescents (DuBay 2007). Some well-known readability formulas that can be used are SMOG (simplified measure of gobbledygook) test of readability (McLaughlin 1969), the Flesch Reading Ease test and the Flesch-Kincaid Grade (Flesch 1948; 1962; Kincaid et al. 1975). The latter readability test can be easily run in various versions of Microsoft Word. de Leeuw (2011) also recommends the Flesch-Kincaid readability formula for general screening purposes in children and adolescent populations.

Format of questions

The most common close-ended format questions used in surveys include dichotomous questions, multiple choice questions, and rating scales (Bradburn et al. 2004; DeVellis 2016). Dichotomous questions (e.g. true/false and agree/disagree) are extremely easy to answer, placing minimal burden on the participant (DeVellis 2016). This format may suit the lower cognitive processes of children and young adolescents (Mellor and Moore 2014). However, a scale consisting of dichotomous questions has lower psychometric properties, compared to a scale with more options (Krosnick and Presser 2010). A major shortcoming of dichotomous questions is that each item can have only minimal variability, which means more items are needed to obtain the same degree of scale variance if the items are dichotomous (DeVellis 2016). If measuring the intensity or extremity of a construct is desirable, dichotomous questions are unsuitable (Alwin 2007; Mitchell and Jolley 2013).

Multiple choice questions (selecting the best possible answer out of specified options) can be answered relatively quickly, and are primary means of student assessment in educational settings (Lissitz and Hou 2012). However, it is believed that multiple choice questions allow for a high probability of guessing correctly which can cause lower response reliabilities, particularly in respondents with lower reading and cognitive ability (Cronbach 1988; Roediger and Marsh 2005). It is also acknowledged that young adolescents are exposed to multiple choice questions at school and this is associated with being tested (Lissitz and Hou 2012; Roediger and Marsh 2005). Therefore, young adolescents may find research surveys with multiple choice format questions less attractive, thus impacting their willingness and motivation to respond, and increasing the likelihood of applying satisficing techniques (Borgers et al. 2000; Krosnick 1991). Survey questions that resemble test items or school questions may result in more socially desirable responses in adolescents, especially the youngest age group (de Leeuw 2011). Multiple choice questions should be used sparingly in young adolescent populations.

There are three common types of rating scales used in surveys: Likert (1932) scaling, semantic differential scales (Osgood et al. 1957), and Thurstone's (1928) scales. Of these, Likert scales are the most universal rating scale format that can be easily understood by participants (DeVellis 2016). In adult and adolescent populations, Likert scales with three or more response options have been found to be more reliable than dichotomous scales (Alwin 2007; Borgers et al. 2004; Krosnick and Presser 2010). Contrary to dichotomous scales, Likert scales provide a measure of intensity, extremity and direction (De Vaus 2014) that give respondents freedom to respond, thus yielding more information than dichotomous items (Mitchell and Jolley 2013). However, like any other closed format questions, Likert scales may suffer from correct guessing, although the probability of successful guessing is less in Likert scales, compared to multiple choice formats (Bradburn et al. 2004). This suggests that Likert scale questions are a suitable choice for adolescents who are more likely to use satisficing techniques.

Likert scales have been reported to be more interesting and reliable than other scales (Robson 2011; Seale 2004), The use of Likert scales when surveying adolescent is therefore recommended to enhance the motivation of respondents to participate, and reduce satisficing techniques (Borgers et al. 2000). Likert scales were originally designed to measure attitudes (Likert, 1932). They have also been successfully used in instruments measuring opinions, knowledge (Alm-Roijer et al. 2004; Michalos et al. 2011; Saffi et al. 2013; Wallace 2009; Zanolin et al. 2007) and in many research projects with children and young adolescents (Harter 2012a; 2012b; McCabe and Ricciardelli 2003; Mellor and Moore 2003; Moore and Mellor 2003). An optimal Likert scale response format should be based on words, as this format has been proved to be the easiest format for children and young adolescents compared to a format based on numbers (Mellor and Moore 2014).

Number of points on Rating Scales and offering a neutral mid-point

Careful consideration needs to be given to the number of response categories within a rating scale because this can have a large effect on data quality (Alwin 2007). In survey research with adult populations, an increase in the number of responses improves data quality (Alwin 2007; Krosnick and Fabrigar 1997; Krosnick and Presser 2010; Lozano et al. 2008; Preston and Colman 2000). However, this phenomenon is inconsistent with young adolescents. Due to developmental differences in communication and reading skills, and poorer memory, too many response options can place a burden on young adolescents, encouraging them to apply a satisficing strategy (Borgers and Hox 2000). Borgers et al. (2004) systematically researched the effects of question characteristics in survey research with children and young adolescents (8 to 16 years) and identified that increasing the number of response options to around seven increased the reliability of responses within the scale. However, increasing the number of response options had a negative effect on the stability of responses over time. In addition, questions with a response scale that offered a neutral midpoint produced a bigger

relative difference over time, especially in girls. Accordingly, it is optimal to offer approximately four response options with young adolescents as respondents.

Researchers have also emphasized that questions offering neutral points will decrease response reliability and decrease measurement quality particularly in children and adolescents (Alwin 2007; Borgers et al. 2000; Borgers and Hox 2000; Krosnick and Fabrigar 1997; Vaillancourt 1973). Furthermore, Mondak and Davis (2001) suggest that when measuring factual knowledge offering a neutral point like a 'don't know' option will lower the validity of responses by masking substantive understanding. Neutral points like 'don't know' discourage respondents to report their true knowledge or opinion (Borgers and Hox 2000; Ogan et al. 2013). Therefore, when developing surveys for adolescents the inclusion of neutral mid-points in survey questions should be avoided.

Verbally labelling response categories

Verbal labelling is using either words or short expressions (e.g. fully, mainly, always, never) to clarify response categories of rating scales (Rohrmann 2007). There is a consensus among researchers that verbally labelling response categories within rating scales can provide a better understanding of each category and clarify its meaning, resulting in lower response bias and more reliable responses (Alwin 2007; Krosnick and Fabrigar 1997; Menold et al. 2014; Moors et al. 2014; Rohrmann 2007). This can be particularly useful for young adolescents as their logical decision making skills are still developing and they are very literal in their interpretation of words (Borgers and Hox 2000; de Leeuw 2011). There are two types of verbal labelling: verbal labels for end categories only (END form), and verbal labels for each of the categories (ALL form). The 'ALL form' labelling has been found to increase the reliability of responses more than 'END form' labelling in adult and adolescent populations (Alwin 2007; Borgers and Hox 2000; de Leeuw 2011; Dillman et al. 2014; Menold et al. 2014; Moors et al. 2014). The results of Borges et al. (2003) experimental research (n: 91, 9-16 years old) also support the use of 'ALL form' labelling,

identifying increases in stability of responses over time in children when using this format, especially for those > 10 years. Therefore, to enhance the reliability of responses in survey research with young adolescents, 'ALL form' labelling should be applied to response categories.

Negative or positive wording

Survey researchers often include the alternation of positively and negatively worded questions to reduce agreement bias in the respondents, thus enhancing the reliability of responses (Cronbach 1950; DeVellis 2016). Agreement bias refers to a respondent's tendency to agree with items, irrespective of their content (DeVellis 2016). However, it is advisable to avoid the use of negatively worded questions in surveys for children and young adolescents as it affects their decision making skills, and increases the cognitive burden and inconsistency in responses, resulting in lower response reliability (Borgers et al. 2000; Borgers and Hox 2000; Borgers et al. 2004; Colosi 2005; de Leeuw 2011). Children, especially those with poorer reading skills, are less able to respond appropriately to negative items and this effect biases the interpretation of their responses (de Leeuw 2011; Marsh 1986). When young adolescents are the target population, survey researchers should avoid utilizing negatively formulated questions.

Survey layout and question sequence

Having a good set of questions to ask the respondent does not guarantee success in conducting a survey. Attention to the survey layout is also necessary to achieve the goals of the survey (De Vaus 2014; Robson 2011). An interesting and clear survey layout can increase the level of motivation and willingness to participate, thus minimizing satisficing techniques (De Vaus 2014). It is also important to keep the survey layout and design consistent as it will help the respondents process information throughout the survey more efficiently (Dillman and Smyth 2007).

Survey researchers need to give special consideration to initial survey instructions, font and spacing to facilitate comprehension and interpretation of the survey (de Leeuw 2011). It has been shown that the number of words in the introductory text within a survey has a positive effect on the reliability of responses in survey research with children and young adolescents; the more words used in the introductory text, the more reliable the responses will be (Borgers and Hox 2000; 2001). Therefore, it is recommended to use a clear and extensive introductory text (words ~ 100) with simple short sentences when surveying children and young adolescents (Borgers and Hox 2000).

The sequence of questions in a survey can also influence the quality of responses and requires consideration. The order of questions needs to be planned, mostly by considering the survey aims (Krosnick and Presser 2010). It is recommended to start the survey with questions which relate to the topic of the research and leave personal, demographic and classificatory questions to the end of survey as respondents may find them irritating or boring (Pole and Lampard 2013). In general, questions in a survey can either be ordered randomly or organized by topic. The purpose of randomizing is to minimize response bias because answers to later questions may be biased by the presentation of earlier questions (Krosnick and Presser 2010). Nevertheless, the completion of a survey consisting of randomized questions places more burden on participants as they need to go back and forth between different topics (Rea and Parker 2005). This can affect motivation, produce fatigue, and increase the likelihood of applying satisficing techniques, especially in young adolescents (Bradburn et al. 2004; de Leeuw 2011; Krosnick and Presser 2010). It is recommended that survey questions have a logical flow, especially when the aim is to assess knowledge (Moser and Kalton 1985; Rea and Parker 2005). Grouping related questions can give survey questions a meaningful flow (Rea and Parker 2005) and help respondents to focus and concentrate on specific issues without distraction (Bradburn et al. 2004; Dillman et al. 2014; Rea and Parker 2005). It is especially useful for young adolescents to facilitate their cognitive processing by specifying the meaning of a question

more clearly, or making retrieval from memory easier, thus improving measurement (de Leeuw 2011; Krosnick and Presser 2010). Therefore, when considering survey structure for adolescent populations, it is advisable to group related questions together to provide optimal flow (De Vaus 2014; Dillman et al. 2014).

When grouping related questions together, it is important to consider the conditioning effect of earlier questions (Dillman et al. 2014; Moser and Kalton 1985). In order to minimize question order bias in adult and adolescent populations, questions with the same topic should be arranged from broad questions to specific issues, using what is known as a 'funnel sequence' (Krosnick and Presser 2010; Moser and Kalton 1985; Seale 2004). In order to reduce the likelihood of successful guessing in questions with the same topic, several questions in different formats focusing on one specific issue should be asked (Bradburn et al. 2004). Additionally, in order for a survey to flow well, and for simplicity and clarity, it is better if response options of survey questions are kept consistent (Moser and Kalton 1985; Rea and Parker 2005).

Grouping related questions with the same topic is optimal in survey research with children and young adolescents when measuring knowledge. However, attitude questions, especially those that ask for general attitude rating, are more susceptible to order effects (Bradburn et al. 2004; Schuman and Presser 1996). The question order effect may be greater in attitude questions when there are related questions with different degrees of specificity. Questions that specify, limit or restrict the way that respondents think about a particular attitude aspect appear to affect the responses to the more general questions related to the attitude being measured (Bradburn et al. 2004). Therefore, survey researchers may prefer to randomly order the questions depending on the aim and purpose of their investigation.

Pre-testing survey with target population

Following guidelines for survey design can enhance survey reliability and validity, however, it is important to consider pre-testing a survey to maximize the success (Dillman and Smyth 2007; Haeger et al. 2012; Krosnick and Presser 2010). It is recommended that survey researchers perform some formal evaluation (pre-testing) prior to administering a survey to their target population to ensure its reliability and validity. The following sections incorporate some important and common pre-testing approaches that survey researchers should undertake prior to administering a survey.

Face and content validity

One of the first steps in the area of survey development is to assess the face and content validity of the survey to make sure that the content assesses what it is supposed to (Bryman 2012; De Vaus 2014; DeVellis 2016). Face validity is simply whether the survey appears (at face value) to measure what it claims to (DeVellis 2016). Content validity is the extent to which the items in a survey reflect or represent the content domain (e.g. the representativeness of the items) (DeVellis 2016; Lawshe 1975; Rossi et al. 2013).

The face validity of a survey should be verified by a panel of subject experts to ensure the items appear to represent adequately the construct under investigation (DeVellis 2016). One-to-one interviews or a focus group can be conducted with a panel of subject experts to check the adequacy of the survey items in assessing the construct under investigation. Items that are considered unsuitable (or irrelevant) and unclear for the purpose of investigation should then be discarded from the survey (Hardesty and Bearden 2004; Nevo 1985).

Content validity assessment should be carried out with both a panel of subject experts and the target population. One way of establishing content validity involves a panel of subject experts considering the importance of individual items within a survey (Bryman 2012; De Vaus 2014; DeVellis 2016; Dillman et al. 2014). This can be established using qualitative methods of data collection (e.g. consultation using one-to-one interviews or focus groups) or using quantitative methods to quantify content validity (Ayre and Scally 2014; DeVellis

2016). A well-known method for quantifying content validity that can be utilized is the Lawshe (1975) method. This involves a panel of subject experts rating each item in a survey into one of three categories: 'essential,' 'useful, but not essential,' or 'not necessary.' Items that are considered 'essential' by a critical number of panel members are then included within the survey, with items failing to achieve this critical level being discarded (Ayre and Scally 2014).

In addition, the target population (in this case, young adolescents) should review survey items to assess their ease of understanding (de Leeuw 2011; DeVellis 2016; Vogt et al. 2004). It is recommended to assess whether questions and response formats are understood consistently across respondents, and in the way intended by researchers (Collins 2003; Willis 2005). This approach is commonly cited as 'cognitive interviewing', which is one of the most prominent methods for identifying and correcting problems with survey questions, especially for improvements in surveys for young adolescents (Beatty and Willis 2007; Collins 2003; de Leeuw 2011; Dillman et al. 2014; Haeger et al. 2012; Presser et al. 2004; Willis 2005). It is recommended to conduct cognitive interviewing in a one-to-one interview or focus group discussions with members of the target population prior to conducting a survey (de Leeuw 2011; Norris et al. 2014; Presser et al. 2004). Two common interview techniques are recommended; think-aloud and verbal probing, for more detailed information, please refer to Beatty and Willis (2000), Haeger et al. (2012) Presser et al. (2004) and Willis (2005).

Item analysis

An item analysis involves statistical analysis that can aid survey researchers to identify items that lack clarity or that may not be appropriate for, or discriminate between, respondents. This is particularly useful when assessing knowledge (Rattray and Jones 2007). The two most common analysis approaches that should be considered in item analysis are evaluation of item difficulty and item discrimination (Kline 2000; Quagrain and Arhin 2017;

Rattray and Jones 2007). Item difficulty is a measure of the percentage of respondents who have answered the item correctly (Kline 2000). High endorsement of a single item is problematic (Priest et al. 1995). Therefore, when $> 80\%$ or $< 20\%$ of respondents answer an item correctly, it is recommended that the item be removed (Kline, 2000). Item discrimination is the ability of an item to discriminate between respondents who do well, and those who do not (Rattray and Jones 2007). An item-total correlation cut-off of < 0.3 can be used to identify items with low discrimination ability, thus informing a decision to remove them (Ferketich 1991; Kline 2000).

Reliability analysis

Reliability refers to the internal consistency, repeatability, or stability of a survey (DeVellis 2016; Rattray and Jones 2007). Two common methods for computing reliability are internal reliability and test-retest reliability (Bryman 2012). Internal consistency reliability is concerned with the homogeneity of the items within a scale and often assessed with Cronbach's coefficient alpha (DeVellis 2016). It is expected that all items correlate well with each other within a scale, and for items to be considered sufficiently reliable, Cronbach's alpha values should be ≥ 0.70 (Bryman 2012; Kline 2000). Test-retest reliability can assess stability of a measure over time (Rattray and Jones 2007; Rossi et al. 2013). Published research has often used correlation (e.g. Pearson's correlation coefficients) for assessing test-retest reliability (Tan et al. 2015). However, correlation is not a suitable method for establishing test-retest reliability due to the fact that agreement and association are two different concepts (Jakobsson and Westergren 2005). Accordingly, perfect association does not mean perfect agreement. In order to establish agreement for ordinal data, Weighted Kappa coefficient is probably the most useful method (Cohen 1968; Jakobsson and Westergren 2005; Tan et al. 2015). Cohen Kappa should be used for establishing test-retest agreement for binary categorical outcomes (Cohen 1960; Sim and Wright 2005). For numeric outcomes with two or more measurements Bland-Altman Limits of agreement and Intraclass Correlation can be used respectively (Bland and Altman 1986; Weir 2005).

Construct Validity

Construct validity relates to how well the items in the survey characterize and measure the underlying construct (Cronbach and Meehl 1955; DeVellis 2016). One approach to investigate construct validity of self-reporting scales is exploratory factor analysis that can be used to determine the domains or constructs within the developing measure (Rattray and Jones 2007; Williams et al. 2010). Another method to investigate construct validity is 'Known-Groups Validation' (DeVellis 2016; Rossi et al. 2013). Known-groups validation typically involves demonstrating that a scale can differentiate across groups that theoretically are expected to be different, based on their scale scores (Cronbach and Meehl 1955; DeVellis 2016; Hattie and Cooksey 1984; Rossi et al. 2013).

Mode of administration

Because of its large impact on data quality, once a survey has been developed and evaluated a decision needs to be made on its mode of administration (Bowling 2005; 2014; Dillman 2006; Dillman et al. 2014). The most commonly used self-administration modes of collecting survey data are paper and pencil, and electronic (computer assisted). Young adolescents tend to use satisficing techniques when they find a survey uninteresting (Borgers et al. 2000). When conducting survey research, especially with young adolescents, online administration is recommended as this mode is more attractive and it enhances motivation of respondents to complete the survey (Borgers et al. 2000; Dillman et al. 2014). This approach also helps to overcome issues related to adolescents increased sensitivity to peer pressure and group norms by providing a more private mode of administration (de Leeuw 2011; van Hattum and de Leeuw 1999).

Research has demonstrated that in comparison with paper and pencil surveys, computer-assisted surveys result in less item non-response and less socially desirable answers when surveying children and young adolescents, improving the reliability of responses (Borgers et al. 2003; van Hattum and de Leeuw 1999). In addition, the use of a computer can help to

avoid some mistakes, such as involuntarily skipping a question, thus less non-response (Bowling 2005; Fargas-Malet et al. 2010; van Hattum and de Leeuw 1999; Dillman et al. 2014).

Limitations in survey development for adolescents' research and suggestions for further research

Data quality in survey research with adolescents is highly dependent on the characteristics and developmental stage of adolescents (Borgers et al. 2000). However, there are many more variables that are associated with the quality of survey data. In a large, longitudinal study with more than 18000 participants aged 10 to 15 years, Maveltova and Lynn (2017) identified lower data quality among young adolescents who live in households with lower socio-economic status compared to those in higher socio-economic status households. Additionally lower data quality was observed when parental expectations and parental engagement in education was low.

Adolescents aged ≥ 11 have most skills needed to complete surveys with minimal errors related to developmental concerns (Arthur et al. 2017). However, not all children have the same opportunities, and even in privileged circumstances, children did not develop equally. Children in less privileged circumstances (lack of malnutrition, schooling, etc.) may develop more slowly (de Leeuw 2011). It is important to consider all these variables when designing and analyzing adolescents survey data. Much remains to be learned about how different variables such as culture, religion, disability and ethnicity affect data quality in survey research with children and adolescents. Research in these areas is necessary to improve survey research quality with children and adolescents.

Conclusion

There are significant developmental differences between adolescents and adult populations. Therefore, adolescents' characteristics should be taken into account in order to develop suitable surveys for this population. Through reviewing a wide range of literature in the area

of survey development and applying satisficing theory and Piaget's theory, a comprehensive and inclusive survey development guide for use with adolescents aged 11 to 16 years has been presented. It is hoped the guide will be useful (and make research less daunting) for those contemplating undertaking research requiring survey development. It is recommended that survey researchers give careful consideration to survey wording and length, and utilize a readability formula to ensure that survey length and wording is suitable for adolescents' age, educational level, and decision making abilities. Closed-format questions are preferable to open-ended questions due to reduced cognitive demands associated with responding to them. Likert scales may be more developmentally appropriate for young adolescents, and, if utilizing ranking scales, a four-point fully labelled scale with no negative wording is optimal. Survey layout and sequence of questions should also be given careful consideration to achieve the goals of the survey research with young adolescents. Prior to administering a survey to the target population, some formal evaluation should be conducted. It is important to establish the face, content and construct validity, as well as the internal and test-re test reliability of a survey, to ensure the survey accuracy and suitability for target population. Lastly, for adolescent populations online administration is the recommended mode of administration. It is envisaged that that these developmental science based recommendations will improve response quality in survey research with young adolescents.

References:

- Alm-Roijer, C., Stagno, M., Uden, G., & Erhardt, L. (2004). Better Knowledge Improves Adherence to Lifestyle Changes and Medication in Patients with Coronary Heart Disease. *European Journal of Cardiovascular Nursing*, 3(4), 321–330. <https://doi.org/10.1016/j.ejcnurse.2004.05.002>
- Alwin, D. F. (2007). *Margins of Error. Wiley Series in Survey Methodology*. John Wiley & Sons, Inc. <https://doi.org/10.1002/9780470146316>
- Arthur, A. M., Smith, M. H., White, A. S., Hawley, L., & Koziol, N. A. (2017). *Age-sensitive instrument design for youth: A developmental approach*. Retrieved from <http://cyfs.unl.edu/resources/downloads/working-papers/MAP-working-paper-2017-1.pdf> (Accessed: 2nd November 2017)

- Ayre, C., & Scally, A. J. (2013). Critical values for Lawshe's content validity ratio revisiting the original methods of calculation. *Measurement and Evaluation in Counseling and Development*, 47(1), 79–86. <https://doi.org/10.1177/0748175613513808>
- Beatty, P. C., Willis, G. B., Beatty, P. C., & Willis, G. B. (2007). The practice of cognitive interviewing. *Public Opinion Quarterly*, 71(2), 287–311. <https://doi.org/10.1093/poq/nfm006>
- Bell, A. (2007). Designing and testing questionnaires for children. *Journal of Research in Nursing*, 12(5), 461–469. <https://doi.org/10.1177/1744987107079616>
- Bland, J. M., & Altman, D. G. (1986). Statistical Methods for Assessing Agreement between Two Methods of Clinical Measurement. *The Lancet*, 327(8476), 307–310. [https://doi.org/10.1016/s0140-6736\(86\)90837-8](https://doi.org/10.1016/s0140-6736(86)90837-8)
- Borgers, N., de Leeuw, E., & Hox, J. J. (2000). Children as Respondents in Survey Research: Cognitive Development and Response Quality. *Bulletin of Sociological Methodology/Bulletin de Methodologie Sociologique*, 66(1), 60–75. <https://doi.org/10.1177/075910630006600106>
- Borgers, N., & Hox, J. J. (2000). Reliability of Responses in Questionnaire Research with Children. In *The Fifth International Conference on Logic and Methodology*. Cologne, Germany.
- Borgers, N., & Hox, J. J. (2001). Item nonresponse in questionnaire research with children. *Journal of Official Statistics*, 17(2), 321–335.
- Borgers, N., Hox, J. J., & Sikkel, D. (2003). Response Quality in Survey Research with Children and Adolescents: The Effect of Labeled Response Options and Vague Quantifiers. *International Journal of Public Opinion Research*, 15(1), 83–94. <https://doi.org/10.1093/ijpor/15.1.83>
- Borgers, N., Hox, J. J., & Sikkel, D. (2004). Response Effects in Surveys on Children and Adolescents: The Effect of Number of Response Options, Negative Wording, and Neutral Mid-Point. *Quality and Quantity*, 38(1), 17–33. <https://doi.org/10.1023/b:ququ.0000013236.29205.a6>
- Bowling, A. (2005). Mode of questionnaire administration can have serious effects on data quality. *Journal of Public Health*, 27(3), 281–291. <https://doi.org/10.1093/pubmed/fdi031>
- Bowling, A. (2014). *Research Methods in Health: Investigating Health and Health Services* (4th ed.). Maidenhead: McGraw-Hill Open University Press.
- Boynton, P. M., & Greenhalgh, T. (2004). Selecting, designing, and developing your questionnaire. *BMJ*, 328(7451), 1312–1315. <https://doi.org/10.1136/bmj.328.7451.1312>
- Bradburn, N. M., Sudman, S., & Wansink, B. (2004). *Asking Questions the Definitive Guide to Questionnaire Design—for Market Research, Political Polls, and Social and Health Questionnaires*. San Francisco: Jossey-Bass Inc., U.S.
- Bryman, A. (2012). *Social Research Methods* (4th ed.). Oxford: University Press.

- Calderon, J. L., Morales, L. S., Liu, H., & Hays, R. (2006). Variation in the Readability of Items Within Surveys. *American Journal of Medical Quality*, 21(1), 49–56. <https://doi.org/10.1177/1062860605283572>
- Cauffman, E., & Steinberg, L. (2000). (Im) maturity of judgment in adolescence: Why adolescents may be less culpable than adults. *Behavioral Sciences & the Law*, 18(6), 741–760.
- Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1), 37–46. <https://doi.org/10.1177/001316446002000104>
- Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70(4), 213–220. <https://doi.org/10.1037/h0026256>
- Collins, D. (2003). Pretesting survey instruments: An overview of cognitive methods. *Quality of Life Research*, 12(3), 229–238. <https://doi.org/10.1023/a:1023254226592>
- Colosi, R. (2005). Negatively worded questions cause respondent confusion. In *U. B. o. t. Census* (pp. 2896–2903). Suitland, MD: ASA Section on Survey Research Methods.
- Cronbach, L. J. (1950). Further Evidence on Response Sets and Test Design. *Educational and Psychological Measurement*, 10(1), 3–31. <https://doi.org/10.1177/001316445001000101>
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281–302. <https://doi.org/10.1037/h0040957>
- Cronbach, L. J. (1988). Five Perspectives on Validity Argument . In H. Wainer & H. I. Braun (Eds.), *Test Validity* (pp. 3–16). New York: Routledge.
- de Leeuw, E., Natacha, B., & Astrid, S. (2004). Pretesting Questionnaires for Children and Adolescents. In S. Presser Rothgeb, J. M., M. P. Couper, J. T. Lessler, E. . Martin, J. Martin, & E. Singer (Eds.), *Methods for Testing and Evaluating Survey Questionnaires* (pp. 409–429). New Jersey: John Wiley & Sons.
- de Leeuw, E. D. (2011). *Improving data quality when surveying children and adolescents: Cognitive and social development and its role in questionnaire construction and pretesting*. In Report prepared for the Annual Meeting of the Academy of Finland: Research Programs Public Health, Finland.
- De Vaus, D. (2014). *Surveys in Social Research*. New York: Routledge.
- Deutskens, E., de Ruyter, K., Wetzels, M., & Oosterveld, P. (2004). Response Rate and Response Quality of Internet-Based Surveys: An Experimental Study. *Marketing Letters*, 15(1), 21–36. <https://doi.org/10.1023/b:mark.0000021968.86465.00>
- DeVellis, R. F. (2016). *Scale development: theory and applications* (4th ed.). London: SAGE.
- Dillman, D. A. (2006). Why Choice of Survey Mode Makes a Difference. *Public Health Reports*, 121(1), 11–13. <https://doi.org/10.1177/003335490612100106>

- Dillman, D. A., & Smyth, J. D. (2007). Design Effects in the Transition to Web-Based Surveys. *American Journal of Preventive Medicine*, 32(5), S90–S96. <https://doi.org/10.1016/j.amepre.2007.03.008>
- Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, Phone Mail and Mixed-Mode Surveys: The Tailored Design Method* (4th ed.). New York: John Wiley and Sons Inc.
- DuBay, W. H. (2007). *Smart Language: Readers, Readability, and the Grading of Text*. Costa Mesa: Booksurge Publishing.
- Fargas-Malet, M., McSherry, D., Larkin, E., & Robinson, C. (2010). research with children: methodological issues and innovative techniques. *Journal of Early Childhood Research*, 8(2), 175–192. <https://doi.org/10.1177/1476718x09345412>
- Feldman, D. H. (2004). Piaget's stages: the unfinished symphony of cognitive development. *New Ideas in Psychology*, 22(3), 175–231.
- Ferketich, S. (1991). Focus on psychometrics. Aspects of item analysis. *Research in Nursing & Health*, 14(2), 165–168. <https://doi.org/10.1002/nur.4770140211>
- Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32(3), 221–233. <https://doi.org/10.1037/h0057532>
- Flesch, R. (1962). *The art of plain talk. Revised edition*. New York: Harper and Row.
- Fuchs, M. (2005). Children and adolescents as respondents. experiments on question order, response order, scale effects and the effect of numeric values associated with response options. *Journal of Official Statistics*, 21(4), 701–725.
- Galesic, M., & Bosnjak, M. (2009). Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public Opinion Quarterly*, 73(2), 349–360. <https://doi.org/10.1093/poq/nfp031>
- Greig, A., Taylor, J., & MacKay, T. (2013). *Doing research with children: a practical guide*. SAGE Publications, Ltd. <https://doi.org/10.4135/9781526402219>
- Haeger, H., Lambert, A. D., Kinzie, J., & Gieser, J. (2012). Using cognitive interviews to improve survey instruments. New Orleans: The Association for Institutional Research Annual Forum.
- Hardesty, D. M., & Bearden, W. O. (2004). The use of expert judges in scale development. *Journal of Business Research*, 57(2), 98–107. [https://doi.org/10.1016/s0148-2963\(01\)00295-8](https://doi.org/10.1016/s0148-2963(01)00295-8)
- Harter, S. (2012a). *Manual for the Self-Perception Profile for Adolescents*. Denver, CO: University of Denver.
- Harter, S. (2012b). *Manual for the Self-Perception Profile for Children*. Denver, CO: University of Denver.
- Hattie, J., & Cooksey, R. W. (1984). Procedures for assessing the validities of tests using the known-groups method. *Applied Psychological Measurement*, 8(3), 295–305.

- Jakobsson, U., & Westergren, A. (2005). Statistical methods for assessing agreement for ordinal data. *Scandinavian Journal of Caring Sciences*. Wiley-Blackwell. <https://doi.org/10.1111/j.1471-6712.2005.00368.x>
- Kelley, K., Clark, B., Brown, V. Sitzia, J. (2003). Good practice in the conduct and reporting of survey research. *International Journal for Quality in Health Care*, 15(3), 261–266. <https://doi.org/10.1093/intqhc/mzg031>
- Kincaid, J. P., Fishburne, R., Rogers, R., & Chissom, B. S. (1975). *Derivation of new readability formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for navy enlisted personnel*. Defense Technical Information Center. <https://doi.org/10.21236/ada006655>
- Kitzinger, J. (1995). Qualitative Research: Introducing focus groups. *BMJ*, 311(7000), 299–302. <https://doi.org/10.1136/bmj.311.7000.299>
- Kline, P. (2000). *The Handbook of Psychological Testing* (2nd ed.). London: Routledge.
- Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5(3), 213–236. <https://doi.org/10.1002/acp.2350050305>
- Krosnick, J. A., & Fabrigar, L. R. (1997). Designing Rating Scales for Effective Measurement in Surveys. In P. B. L. Lyberg M. Collins, E. de Leeuw, C. Dippo, N. Schwarz, and D. Trewin (Ed.), *Survey Measurement and Process Quality* (pp. 141–164). New York: Wiley.
- Krosnick, J. A., & Presser, S. (2010). Question and Questionnaire Design. In J. D. W. and P. V. Marsden (Ed.), *Handbook of Survey Research* (pp. 263–314). Bingley: Emerald Group Publishing Ltd.
- Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28(4), 563–575. <https://doi.org/10.1111/j.1744-6570.1975.tb01393.x>
- Lenzner, T. (2014). Are Readability Formulas Valid Tools for Assessing Survey Question Difficulty? *Sociological Methods and Research*, 43(4), 677–698. <https://doi.org/10.1177/0049124113513436>
- Leshem, R. (2016). Brain development, impulsivity, risky decision making, and cognitive control: integrating cognitive and socioemotional processes during adolescence—an introduction to the special issue. *Developmental Neuropsychology*, 41(1–2), 1–5. <https://doi.org/10.1080/87565641.2016.1187033>
- Likert, R. (1932). A Technique for the Measurement of Attitudes. *Archives of Psychology*, 22(140), 1–55.
- Lissitz, R. W., & Hou, X. (2012). Multiple choice items and constructed response items: does it matter? Retrieved from [http://www.education.umd.edu/MARC/multiple choice items and constructed response items.pdf](http://www.education.umd.edu/MARC/multiple%20choice%20items%20and%20constructed%20response%20items.pdf) (Accessed:25th June, 2016)
- Lozano, L.M., García-Cueto, E. & Muñiz, J. (2008). Effect of the number of response categories on the reliability and validity of rating scales. *Methodology*, 4(2), 73–79. <https://doi.org/10.1027/1614-2241.4.2.73>

- Marcus, B., Bosnjak, M., Lindner, S., Pilischenko, S., & Schütz, A. (2007). Compensating for low topic interest and long surveys. *Social Science Computer Review*, 25(3), 372–383. <https://doi.org/10.1177/0894439307297606>
- Marsh, H. W. (1986). Negative item bias in ratings scales for preadolescent children: A cognitive-developmental phenomenon. *Developmental Psychology*, 22(1), 37–49. <https://doi.org/10.1037//0012-1649.22.1.37>
- Mavletova, A., & Lynn, P. (2017). *Data quality in the understanding society youth self-completion questionnaire. Understanding Society at the Institute for Social and Economic Research* (No. 2017 – 8).
- McCabe, M. P., & Ricciardelli, L. A. (2003). Body image and strategies to lose weight and increase muscle among boys and girls. *Health Psychology*, 22(1), 39–46. <https://doi.org/10.1037//0278-6133.22.1.39>
- McLaughlin, G. H. (1969). SMOG grading: A New Readability Formula. *Journal of Reading*, 12(8), 639–646.
- Mellor, D., & Moore, K. A. (2003). The questionnaire on teacher interaction: assessing information transfer in single and multi-teacher environments. *Journal of Classroom Interaction*, 38(2), 29–35.
- Mellor, D., & Moore, K. A. (2014). The use of Likert scales with children. *Journal of Pediatric Psychology*, 39(3), 369–379. <https://doi.org/10.1093/jpepsy/jst079>
- Menold, N., Kaczmirek, L., Lenzner, T., & Neusar, A. (2014). How do respondents attend to verbal labels in rating scales? *Field Methods*, 26(1), 21–39. <https://doi.org/10.1177/1525822x13508270>
- Michalos, A. C., Creech, H., Swayze, N., Maurine Kahlke, P., Buckler, C., & Rempel, K. (2011). Measuring knowledge, attitudes and behaviours concerning sustainable development among tenth grade students in Manitoba. *Social Indicators Research*, 106(2), 213–238. <https://doi.org/10.1007/s11205-011-9809-6>
- Mitchell, M. L., & Jolley, J. M. (2013). *Research Design Explained* (8th ed.). Boston: Wadsworth Publishing Co.
- Mondak, J. J., & Davis, B. C. (2001). Asked and answered: knowledge levels when we won't take "don't know" for an answer. *Political Behavior*, 23(3), 199–224. <https://doi.org/10.1023/a:1015015227594>
- Moore, K. A., & Mellor, D. J. (2003). The nature of children's social interactions at school. *School Psychology International*, 24(3), 329–339. <https://doi.org/10.1177/01430343030243005>
- Moors, G., Kieruj, N. D., & Vermunt, J. K. (2014). The effect of labeling and numbering of response scales on the likelihood of response bias. *Sociological Methodology*, 44(1), 369–399. <https://doi.org/10.1177/0081175013516114>
- Morgan, D. L. (1996). Focus groups. *Annual Review of Sociology*, 22(1), 129–152. <https://doi.org/10.1146/annurev.soc.22.1.129>

- Moser, C., & Kalton, G. (1985). *Survey Methods in Social Investigation* (2nd ed.). Aldershot: Gower.
- Nevo, B. (1985). Face validity revisited. *Journal of Educational Measurement*, 22(4), 287–293. <https://doi.org/10.1111/j.1745-3984.1985.tb01065.x>
- Norris, A. E., Torres-Thomas, S., & Williams, E. T. (2014). Adapting cognitive interviewing for early adolescent hispanic girls and sensitive topics. *Hispanic Health Care International*, 12(3), 111–119. <https://doi.org/10.1891/1540-4153.12.3.111>
- Ogan, C., Karakuay, T., & Kurayun, E. (2013). Methodological issues in a survey of children's online risk-taking and other behaviours in Europe. *Journal of Children and Media*, 7(1), 133–150. <https://doi.org/10.1080/17482798.2012.739812>
- Osgood, C. E. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The Measurement of Meaning*. United States: University of Illinois Press.
- Partridge, B. C. (2010). Adolescent psychological development, parenting styles, and pediatric decision making. *Journal of Medicine and Philosophy*, 35(5), 518–525.
- Piaget, J. (1929). *Child's Conception of the World (International Library of Psychology)*. London: Routledge and Kegan Paul.
- Pole, C., & Lampard, R. (2013). *Practical Social Investigation. Qualitative and Quantitative Methods in Social Research*. Harlow: Printice Hall.
- Presser, S., Couper, M. P., Lessler, J. T., Martin, E., Martin, J., Rothgeb, J. M., & Singer, E. (2004). Methods for testing and evaluating survey questions. *Public Opinion Quarterly*, 68(1), 109–130. <https://doi.org/10.1093/poq/nfh008>
- Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, 104(1), 1–15. [https://doi.org/10.1016/s0001-6918\(99\)00050-5](https://doi.org/10.1016/s0001-6918(99)00050-5)
- Priest, J., Thomas, L., & Bond, S. (1995). Developing and refining a new measurement tool. *Nurse Researcher*, 2(4), 69–81. <https://doi.org/10.7748/nr.2.4.69.s8>
- Quaigrain, K., & Arhin, A. K. (2017). Using reliability and item analysis to evaluate a teacher-developed test in educational measurement and evaluation. *Cogent Education*, 4(1). <https://doi.org/10.1080/2331186x.2017.1301013>
- Rattray, J., & Jones, M. C. (2007). Essential elements of questionnaire design and development. *Journal of Clinical Nursing*, 16(2), 234–243. <https://doi.org/10.1111/j.1365-2702.2006.01573.x>
- Rea, L. M., & Parker, R. A. (2005). *Designing and Conducting Survey Research: A Comprehensive Guide* (3rd ed.). San Francisco: Jossey-Bass.
- Reja, U., Manfreda, K. L., Hlebec, V., & Vehovar, V. (2003). Open-ended vs. close-ended questions in web questionnaires. *Developments in Applied Statistics*, 19, 159–177.
- Reyna, V. F., & Rivers, S. E. (2008). Current theories of risk and rational decision making. *Developmental Review*, 28(1), 1–11.

- Robson, C. (2011). *Real World Research: A Resource for Users of Social Research Methods in Applied Settings* (3rd ed.). Chichester: Wiley.
- Roediger, H. L., & Marsh, E. J. (2005). The positive and negative consequences of multiple-choice testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(5), 1155–1159. <https://doi.org/10.1037/0278-7393.31.5.1155>
- Rohrmann, B. (2007). *Verbal qualifiers for rating scales: sociolinguistic considerations and psychometric data*. University of Melbourne/Australia. Retrieved from <http://www.rohrmannresearch.net/pdfs/rohrmann-vqs-report.pdf> (Accessed: 15th April 2016)
- Rossi, P. H., Wright, J. D., & Anderson, A. B. (2013). *Handbook of survey research*. New York: Academic Press.
- Saffi, M. A. L., Macedo Junior Jacques de, L. J., Trojahn, M. M., Polanczyk, C. A., & Rabelo-Silva, E. (2013). Validity and reliability of a questionnaire on knowledge of cardiovascular risk factors for use in Brazil. *Revista Da Escola de Enfermagem Da USP*, 47(5), 1083–1089. <https://doi.org/10.1590/s0080-623420130000500011>
- Schuman, H., & Presser, S. (1996). *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context*. New York: Academic Press.
- Scott, E. S., & Steinberg, L. (2008). Adolescent development and the regulation of youth crime. *The Future of Child*, 18(2), 15–33. <https://doi.org/10.1353/foc.0.0011>
- Scott, J. (1997). Children as Respondents: Methods for Improving Data Quality. In P. B. L. Lyberg M. Collins, E. de Leeuw, C. Dippo, N. Schwarz, and D. Trewin (Ed.), *Survey Measurement and Process Quality* (pp. 331–350). New York: Wiley-Blackwell.
- Seale, C. (2004). *Social research methods: A reader*. London: Routledge.
- Sim, J., & Wright, C. C. (2005). The Kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Physical Therapy*, 85(3), 257–268. <https://doi.org/10.1093/ptj/85.3.257>
- Smith, K., & Platt, L. (2013). *How Do Children Answer Questions About Frequencies and Quantities? Evidence from A Large-Scale Field Test*. Centre for Longitudinal Studies, Institute of Education, University of London.
- Somerville, L. H. (2013). The teenage brain: sensitivity to social evaluation. *Current Directions in Psychological Science*, 22(2), 121–127. <https://doi.org/10.1177/0963721413476512>
- Tan, C. L., Hassali, M. A., Saleem, F., Shafie, A. A., Aljadhey, H., & Gan, V. B. (2015). Development, test-retest reliability and validity of the Pharmacy Value-Added Services Questionnaire (PVASQ). *Pharmacy Practice*, 13(3), 598. <https://doi.org/10.18549/pharmpract.2015.03.598>
- Thurstone, L. L. (1928). Attitudes Can Be Measured. *American Journal of Sociology*, 33(4), 529–554. <https://doi.org/10.1086/214483>

- Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The Psychology of Survey Response*. Cambridge University Press. <https://doi.org/10.1017/cbo9780511819322>
- Turner, J. (1975). *Cognitive Development*. London: Methue.
- Vaillancourt, P. M. (1973). Stability of children's survey responses. *Public Opinion Quarterly*, 37(3), 373. <https://doi.org/10.1086/268099>
- van Hattum, M. J. C., & de Leeuw, E. (1999). A disk by mail survey of pupils in primary schools: data quality and logistics. *Journal of Official Statistics*, 15(3), 413–429.
- Vogt, D. S., King, D. W., & King, L. A. (2004). Focus Groups in Psychological Assessment: Enhancing Content Validity by Consulting Members of the Target Population. *Psychological Assessment*, 16(3), 231–243. <https://doi.org/10.1037/1040-3590.16.3.231>
- Wallace, C. M. (2009). *Measuring Changes in Attitude, Skill and Knowledge of Undergraduate Nursing Students after Receiving an Educational Intervention in Intimate Partner Violence*.
- Weir, J. P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *Journal of Strength and Conditioning Research*, 19(1), 231–240. <https://doi.org/10.1519/00124278-200502000-00038>
- Williams, B., Onsmann, A., & Brown, T. (2010). Exploratory factor analysis: A five-step guide for novices. *Australasian Journal of Paramedicine*, 8(3), 1–13.
- Willis, G. G. B. (2005). *Cognitive Interviewing: A Tool for Improving Questionnaire Design. PsycEXTRA Dataset*. Thousand Oaks, CA: SAGE Publications.
- Zanolin, M. E., Visentin, M., Trentin, L., Saiani, L., Brugnolli, A., & Grassi, M. (2007). A Questionnaire to Evaluate the Knowledge and Attitudes of Health Care Providers on Pain. *Journal of Pain and Symptom Management*, 33(6), 727–736. <https://doi.org/10.1016/j.jpainsymman.2006.09.032>

Table 1: Key considerations for survey development for young adolescents

<ul style="list-style-type: none"> ▪ Question characteristics affect the reliability of responses in survey research with young adolescents.
<ul style="list-style-type: none"> ▪ Surveys for young adolescents should be developed considering their characteristics and developmental stages.
<ul style="list-style-type: none"> ▪ Optimal survey length for young adolescents is ~ 10 minutes.
<ul style="list-style-type: none"> ▪ A readability formula should be utilised to ensure the survey wording and length is suitable for young adolescents.
<ul style="list-style-type: none"> ▪ Most suitable closed format scale for young adolescents is a fully labelled Likert scale with four response options, preferably with no negative wording.
<ul style="list-style-type: none"> ▪ A clear introductory text around ~ 100 words should be provided when surveying children and young

adolescents.
▪ Grouping questions with the same topic together is the optimal question order for young adolescents.
▪ Prior to administrating a survey to young adolescents, it should be formally pre-tested to ensure its reliability and validity.
▪ Online surveys are more attractive for young adolescents and will enhance the response rate.

In Press