Rates of Missing Responses in Personal Digital Assistant (PDA) Versus Paper Assessments
Lori-Ann Palen, John W. Graham, Edward A. Smith, Linda L. Caldwell, Catherine Mathews and Alan J. Flisher
*Eval Rev* 2008 32: 257
DOI: 10.1177/0193841X07307829

The online version of this article can be found at:
http://erx.sagepub.com/content/32/3/257

Additional services and information for *Evaluation Review* can be found at:

**Email Alerts:** http://erx.sagepub.com/cgi/alerts

**Subscriptions:** http://erx.sagepub.com/subscriptions

**Reprints:** http://www.sagepub.com/journalsReprints.nav

**Permissions:** http://www.sagepub.com/journalsPermissions.nav

**Citations:** http://erx.sagepub.com/content/32/3/257.refs.html
Rates of Missing Responses in Personal Digital Assistant (PDA) Versus Paper Assessments

Lori-Ann Palen
John W. Graham
Edward A. Smith
Linda L. Caldwell
*The Pennsylvania State University*

Catherine Mathews
*Medical Research Council (South Africa)*

Alan J. Flisher
*University of Cape Town*

This article describes rates of missing item responses in personal digital assistant (PDA) assessments as compared to paper assessments. Data come from the evaluation of a classroom-based leisure, life skills, and sexuality education program delivered to high school students in Cape Town, South Africa. Analyses show that the paper assessments had much higher rates of missingness than PDA assessments. This association is moderated by item order. Certain analyses also suggest that paper assessments have higher rates of missingness for items pertaining to participants’ sexual behavior. Implications of these results for evaluation research will be discussed.

**Keywords:** program evaluation; methodology; electronic assessment; missing data

Computerized assessments have been used in social science and health research for more than two decades (Nicholls and Groves 1986). Handheld computers (also called personal digital assistants [PDAs]) have come into use more recently, reducing the cost and increasing the ease of assessments. PDAs have been used in studies covering a wide range of topics and populations, from field worker reports on alcohol at community festivals (Fletcher...
et al. 2003) to college students’ daily diary reports of affect (Green et al. 2006) to performance assessments of medical and dental students (Treadwell 2006).

The advantages of using PDAs are numerous, for both respondents and researchers. Respondents find PDAs to be user friendly, comfortable, and enjoyable (Fletcher et al. 2003; Treadwell 2006). When asked directly which format they prefer, there is evidence that student respondents prefer PDAs to paper (Zwarenstein et al. 2006). As an added benefit to respondents, PDA-based assessments require less time to complete than paper assessments (Forster et al. 1991; Zwarenstein et al. 2006).

PDAs, like other forms of computerized assessment, offer several logistical advantages for researchers. PDA-based assessments can be very flexible, accommodating multilingual populations and complicated skip patterns (Zwarenstein et al. 2006). For example, given proper programming, the same set of PDAs could be brought to any classroom in a South African data collection, regardless of whether the students to be measured on that particular day spoke English or Afrikaans. At the beginning of the assessment, students simply indicate which language they speak, and the correct translation is presented. This is contrasted with multilingual paper-based assessments, in which different versions of a survey need to be brought on each data collection, and the correct translation must be distributed on a person-by-person basis.

There are also many areas of research in which investigators want to present a subset of items only to participants meeting certain criteria. For example, a researcher may be interested in measuring level of condom use among participants who are sexually active. Responses from those who are not sexually active are not of interest (nor even meaningful). Paper assessments can handle these types of items in several ways. There can be a lead question that sorts participants into groups (e.g., sexually active and not) and then instructs certain groups to skip past follow-up items. Alternatively, follow-up items could include a “not applicable” response option. However, either of these survey design strategies has the potential to be confusing to participants. PDAs can simplify the assessment process by only presenting follow-up items to subgroups of interest.

There are several additional advantages of using PDAs in research. Data from PDA assessments are available for analysis almost immediately following their collection (Fletcher et al. 2003; Forster et al. 1991; Tapp et al.

Authors’ Note: This research was funded by NIH Grants R01 DA01749 and T32 DA017629–01A1. The authors wish to thank John Dziak for his assistance with the revision of this article. Correspondence should be addressed to Lori-Ann Palen, Department of Human Development & Family Studies, S-113 Henderson Building, University Park, PA 16802; phone: (814) 865-5206; fax: (814) 865-2530; e-mail: lxp201@psu.edu.
2006; Treadwell 2006; Zwarenstein et al. 2006). In addition, although the set-up costs of PDA assessments are higher than paper, there are benefits of economies of scale across repeated assessments (Tapp et al. 2006; Zwarenstein et al. 2006).

Available evidence on the comparative quality of PDA data suggests that, at worst, their quality is equal to that of paper assessments. Numerous studies (e.g., Fletcher et al. 2003; Green et al. 2006; Tapp et al. 2006; Treadwell 2006; Zwarenstein et al. 2006) have shown that PDAs and paper yield comparable data, even in assessments of risk behaviors such as substance use and certain sexual behaviors (Bernhardt, Usdan, and Burnett 2005; Jaspan et al. 2007; LaBrie et al. 2006). There is also some evidence that PDAs yield better quality data than paper assessments. For example, Forster et al. (1991) found that fieldworker data on malaria morbidity in Gambia was more accurate for PDA assessments than paper.

We are aware of only a few studies that have examined rates of missingness in PDA data. In one study using adult field workers for observational data collection, PDAs resulted in less missing data than paper assessments for close-ended questions (Fletcher et al. 2003). In a diary study of physical pain in children, those who used electronic diaries completed both more days of assessment and more items within assessments than those who used paper diaries (Palermo, Valenzuela, and Stork 2004). However, it is important to note that each of the studies were small enough that respondents could receive fairly individualized or intense training, assistance, and supervision.

A larger study ($N = 212$) of South African adolescents showed that a paper-based assessment yielded 7 times more missing values than an identical PDA assessment given to the same youth 2 weeks later (Jaspan et al. 2007). It is unclear whether this association will be as strong in a large-scale evaluation study of adolescents. Also, it is not clear whether PDA assessments are subject to the same type of “drop-off” seen in long paper surveys, in which items closer to the end of a survey are more likely to be left blank because of factors such as slow reading speed (Graham, Cumsille, and Elek-Fisk 2003) or participant fatigue.

In addition, there is uncertainty as to whether patterns of missingness will be different for questions of a sensitive nature, such as those pertaining to sexual behavior or substance use. In one study, students stated that they felt that electronic assessments were more confidential and made it easier to give truthful answers about sexual behavior (Zwarenstein et al. 2006). However, another study failed to find these clear associations (Jaspan et al. 2007). We are not aware of any studies that have directly tested whether the sensitivity of the questions has an impact on response rate for PDAs.
The Current Study

The present study sought to describe rates of missing item responses in PDA assessments, as compared to paper assessments. It also examined whether any differences in missingness were moderated by the order of items presented or whether the items pertained to sexual behavior or substance use.

Method

Sample

The participants in this study consisted of two cohorts of eighth-grade students from Mitchell’s Plain, a low-income township in Cape Town, South Africa. These students were participating in a randomized control trial of a classroom-based leisure, life skills, and sexuality education program (Caldwell et al. 2004). In determining which of 25 Mitchell’s Plain high schools to include in the trial, 6 were excluded because of concerns about their ability to functionally participate. Of the remaining schools, 4 were randomly selected to receive the HealthWise curriculum, and 5 schools were subjectively matched to the treatment schools to serve as controls.

In 2003, 345 students participated in a paper-based posttest assessment, and 1,980 students participated in an electronic posttest assessment the following year. Basic demographic information for these samples appears in Table 1. The two samples were similar in age and gender composition, but there were significant differences in racial composition, with students in the PDA sample more likely to be colored and less likely to be Black than the paper-based sample.

Procedure

Consent/assent. Prior to data collection, an information letter describing the study was mailed to the parents of eligible students. Parents were asked to complete and return an enclosed form if they did not want their child to participate in the study. At the time of the data collection, any student for whom parental consent was not refused was asked to provide assent.

Paper assessments. In 2003, as part of the pilot phase of the intervention trial, two classrooms were randomly selected from all of the intervention schools and four of the control schools. To simplify the development of a pilot curriculum, the intervention sample was limited to English-speaking students.
However, control group participants spoke either English or Afrikaans. Students in the selected classrooms completed a posttest survey near the end of their eighth-grade school year. This survey consisted of 195 questions on topics including leisure activities, sexual behavior, and substance use.

Electronic assessments. In 2004, the first cohort of eighth-grade students began participating in the full intervention trial. Participants completed semi-annual assessments (beginning and end of each school year) on PDAs. Surveys were completed in the students’ home language, either English or Afrikaans. Although not identical to the paper assessment, the PDA assessment also consisted of 201 questions about leisure, sex, and substances. (We chose to omit four additional free-response items from the present analysis, given that there was no analogue to this question format in the paper assessment.) About 8% of eligible students did not complete the survey because of repeated school absence. Three percent of eligible students did not complete the survey because of a refusal of parental consent or student assent.

For data to be developmentally comparable across assessment modes, we used data from the posttest PDA assessment (end of eighth grade) for the present study.

Table 1
Participant Demographics

<table>
<thead>
<tr>
<th></th>
<th>Paper Assessment (N = 345)</th>
<th>PDA Assessment (N = 1,980)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>%</td>
<td>M</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>42.3</td>
<td>48.3</td>
</tr>
<tr>
<td>Female</td>
<td>57.7</td>
<td>51.7</td>
</tr>
<tr>
<td>Racea</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colored</td>
<td>72.8</td>
<td>85.8</td>
</tr>
<tr>
<td>Black</td>
<td>23.4</td>
<td>8.8</td>
</tr>
<tr>
<td>White</td>
<td>1.8</td>
<td>3.9</td>
</tr>
<tr>
<td>Indian</td>
<td>1.2</td>
<td>0.6</td>
</tr>
<tr>
<td>Asian</td>
<td>0.9</td>
<td>—</td>
</tr>
<tr>
<td>Other</td>
<td>—</td>
<td>0.8</td>
</tr>
<tr>
<td>Age (years)</td>
<td>14.0</td>
<td>14.0</td>
</tr>
</tbody>
</table>

Note: PDA = personal digital assistant.

a. Racial composition of the two samples was significantly different, $\chi^2(4) = 65.9, p < .001$. 

However, control group participants spoke either English or Afrikaans.
Measures

Missing item responses. For participants in the paper assessment, a response was considered missing if it was left completely blank. For participants in the PDA assessment, items were presented one at a time, and respondents were not permitted to move on to the next item until the current item was answered. Therefore, it was impossible to have a blank electronic response. However, each item had a “refuse to answer” option, which was considered to be missing for the present study.

Given skip patterns built into the questionnaires, not all participants were presented with all items. In the paper survey, participants who indicated that they had never had sexual intercourse were instructed to skip all follow-up questions related to sexual behavior. In the case of the PDA assessment, participants who answered “no” or “none” to initial questions about leisure, substance use, or sex (e.g., “During the past four weeks have you spent time doing volunteer work?” “How many drinks of alcohol have you had in your entire life?”) were not presented with follow-up questions related to that specific behavior. Participants who refused to answer the lead questions in the PDA assessment were presented with all of the related follow-up questions.

For each item, we calculated the percentage of students who were missing a response to that item. Given the skip patterns, certain missing responses to follow-up items were legitimate. We initially chose to retain these items in the analyses but exclude participants who were missing this response because they were ineligible to answer the item. Therefore, the calculation of a percentage missingness because of omission (rather than a legitimate skip) required a determination of how many students were eligible to respond to each follow-up question.

For the bulk of our analyses, we considered “eligible” students to be those who responded affirmatively to the lead-in question (e.g., a participant who answered “yes” to the item about lifetime sexual intercourse would be eligible to answer a follow-up item about condom use). However, we also conducted additional analyses in which students with missing responses to the lead-in question were also considered eligible to receive the follow-up question. We also considered the possibility that, given differences in the ways that skip patterns are handled in paper and PDA assessments, missingness for follow-up items was not directly comparable between the two assessment formats. Therefore, we also conducted analyses in which follow-up items were completely excluded.

Item characteristics. Each item was assigned a number (1 to 201), corresponding to the order in which it was presented in its respective survey. Each
item was also coded as to whether it pertained to a respondent’s own sexual or substance use behaviors (e.g., “How many sexual partners have you had in the last 6 months?” “How many cigarettes have you smoked in your entire life?”). Items that pertained to both sex and substance use (e.g., “The last time you had sex, did you drink alcohol?”) were coded as having both characteristics. The PDA assessment had 22 sex items and 16 substance use items, whereas the paper assessment had 29 and 17, respectively.

### Analytic Strategy

The unit of analysis was individual items. We used a series of linear regression models to test the associations between percentage of missing responses and assessment format, item order, and item content (whether it was related to sexual or substance use behavior). An initial model was tested that included the control variables of item order, sexual content, and substance use content. Assessment format was added to a subsequent model to test its contribution above and beyond these item-level control variables. Then, each of three interactions between assessment format and item characteristics (item order, sexual behavior item, and substance use item) were added to the model.

### Results

Among all participants, there were demographic group differences in rates of missingness (see Table 2). Boys were more likely to have missing responses than girls. In terms of racial differences, Black students were most likely to have missing responses and Indian students were least likely. Older students were more likely to have missing responses than younger students, and students who took the PDA survey in Afrikaans were more likely to have missing responses than students who took the PDA survey in English.

In the PDA assessment, on average, items were missing 0.8% ($SD = .51$) of responses. Items in the paper assessment were missing an average of 11.5% ($SD = 6.48$) of responses. Table 3 displays the percentage of missing responses for different types of items.

The results of linear regression models predicting percentage of missing responses appear in Table 4. Model 1 includes the three item-level characteristics examined in this study. In this model, item order was the only significant predictor of percentage missingness, with later items missing more responses. Missingness was not significantly different for items that were or were not about sex or substance use.
Model 2 shows the added contribution of survey format (PDA vs. paper) to the prediction of percentage missingness. Survey format was a significant predictor of missingness, increasing the explanatory power of the model by nearly 6 times.

Model 3 adds the test for an interaction between assessment format and item order. This interaction was statistically significant, indicating that the increase in missingness over the length of the survey was significantly greater for the paper survey. Figure 1 shows a plot of missingness by item order for each of the two formats.

### Table 2

Poisson Regression Estimates for Demographics Predicting Count of Missing Responses\(^a\)

<table>
<thead>
<tr>
<th>Demographic Variable</th>
<th>( N )</th>
<th>Estimate(^b)</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender (0 = boy, 1 = girl)</td>
<td>2,263</td>
<td>-0.44</td>
<td>.02</td>
</tr>
<tr>
<td>Race(^c)</td>
<td>2,260</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td></td>
<td>1.07</td>
<td>.02</td>
</tr>
<tr>
<td>White</td>
<td></td>
<td>0.47</td>
<td>.07</td>
</tr>
<tr>
<td>Indian</td>
<td></td>
<td>-0.76</td>
<td>.18</td>
</tr>
<tr>
<td>Asian/other</td>
<td></td>
<td>0.32</td>
<td>.12</td>
</tr>
<tr>
<td>Age</td>
<td>2,297</td>
<td>0.22</td>
<td>.01</td>
</tr>
<tr>
<td>Language (0 = English, 1 = Afrikaans)</td>
<td>1,980(^d)</td>
<td>0.77</td>
<td>.04</td>
</tr>
</tbody>
</table>

\(^a\) Each demographic variable was entered in a separate model. Only items asked of all participants were included in counts; skip pattern items were omitted. Given that the number of potential missing responses varied slightly by survey format (149 personal digital assistant [PDA], 167 paper), format was controlled in all models.

\(^b\) All estimates were significant at \( p < .01 \) or less.

\(^c\) Colored students were the reference group.

\(^d\) Data on language were only available for the PDA assessment.

### Table 3

Percentage of Missing Responses by Item Type

<table>
<thead>
<tr>
<th></th>
<th>Paper assessment</th>
<th>PDA assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>All items (mean)</td>
<td>11.5</td>
<td>0.8</td>
</tr>
<tr>
<td>First item</td>
<td>0.9</td>
<td>1.4</td>
</tr>
<tr>
<td>Last item</td>
<td>21.7</td>
<td>1.5</td>
</tr>
<tr>
<td>Sex items (mean)</td>
<td>9.2</td>
<td>1.1</td>
</tr>
<tr>
<td>Substance items (mean)</td>
<td>4.4</td>
<td>0.4</td>
</tr>
</tbody>
</table>

Note: PDA = personal digital assistant. All figures are percentages.
To determine if the remaining interactions contributed to missingness over and above the effect of the format by item order interaction, we included the latter in all subsequent models. We tested two models in which an interaction between assessment format and either sexual or substance use content was included. Neither term reached significance (results not shown). To determine the explanatory power of all predictors of interest, we estimated a final model in which all three interactions were included simultaneously (Model 4; Table 4). The $R^2$ statistic for this model was 0.78.2

We also explored how alternative strategies for dealing with this missingness issue might affect our statistical conclusions. We reestimated Model 4 in two different ways (see Table 5). First, we recalculated percent missingness so that the pool of participants who were eligible for any follow-up questions included not only those who responded affirmatively to the lead question (as in our original analyses) but also those who were missing on the lead question. In this model, both the main effect for sexual content and the interaction between sexual content and assessment format were significant. This indicated that when missingness for a lead question is taken into account, participants were missing more data for sexually related items than for other types and that this association was particularly strong for participants who completed paper assessments.
The second follow-up analysis only included items that were presented to all participants; 51 and 28 follow-up items were excluded for the PDA and paper assessments, respectively. In this model, the assessment format by sexual item content interaction was significant, again indicating that participants in the paper assessment were especially likely to leave the lead sexual behavior item blank, as compared to items for other content areas.

### Discussion

We have demonstrated that even after controlling for other item characteristics, missingness is much higher for paper assessments than for assessments using PDAs. All else being equal, our estimates were that on average, an item in a paper assessment will be missing 11% more responses than an item in a PDA assessment. It appears that this effect is partially driven by the fact that participants in the paper assessment increased sharply in missing responses over the length of the survey whereas participants in the PDA assessment tended to have more consistent, low rates of missingness. Although this phenomenon may be of little consequence for brief survey instruments, it becomes more problematic with lengthier surveys. For

![Table 4](https://example.com/table4.png)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$B$</td>
<td>$SE_B$</td>
<td>$B$</td>
<td>$SE_B$</td>
</tr>
<tr>
<td>Item order</td>
<td>0.04***</td>
<td>0.01</td>
<td>0.04***</td>
<td>.00</td>
</tr>
<tr>
<td>Item about own sexual behavior</td>
<td>1.60</td>
<td>1.05</td>
<td>0.71</td>
<td>.62</td>
</tr>
<tr>
<td>Item about own substance use</td>
<td>-2.38</td>
<td>1.26</td>
<td>-2.46**</td>
<td>.74</td>
</tr>
<tr>
<td>Format ($0 = PDA$, $1 = paper$)</td>
<td>10.79***</td>
<td>.40</td>
<td>10.77***</td>
<td>.34</td>
</tr>
<tr>
<td>Format × item order</td>
<td>0.07***</td>
<td>.01</td>
<td>0.07***</td>
<td>.01</td>
</tr>
<tr>
<td>Format × sex item</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Format × substance use item</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>.10***</td>
<td>.69***</td>
<td>.78***</td>
<td>.78***</td>
</tr>
</tbody>
</table>

Note: PDA = personal digital assistant.  
*p < .05. **p < .01. ***p < .001.
example, if one were to administer a 10-question survey, we estimate that rates of missing responses would increase about 1% between the first and last item, regardless of survey format. If the instrument were instead 100 questions long, we estimate that there would be 4% more missing responses for the first item than the last item. If assessment format is taken into account with this longer survey, we estimate that there would be an additional 7% of responses (for a total of 11%) missing for the final item in a paper-based assessment.

There are several potential explanations for the larger item order effect in the paper assessments. First, given the logistics of the PDA software, an assessment cannot be completed (and subsequent assessments begun on the device) until all items have been completed. Therefore, as compared to paper assessments, it may be both easier and more important for fieldworkers to check whether a survey has been completed. This may result in more encouragement for participants to turn in complete surveys and the allowance of sufficient time for survey completion. Second, the higher novelty of and preference for PDA assessments may increase adolescents’ motivation to complete the assessment. Also, if it takes longer to complete a survey in paper than on a PDA, participants in the paper-based assessment may be especially susceptible to fatigue as they proceed through the questionnaire.

Table 5
Model 4 With Alternative Parameterizations of Participant and Item Eligibility

<table>
<thead>
<tr>
<th>Model With Participants Who Are Affirmative or Missing for Lead Item as Eligible for Follow-Up Item</th>
<th>Model Without Follow-Up Items Included</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
<td>B</td>
</tr>
<tr>
<td>Item order</td>
<td>0.04***</td>
</tr>
<tr>
<td>Item about own sexual behavior</td>
<td>18.34***</td>
</tr>
<tr>
<td>Item about own substance use</td>
<td>–0.60</td>
</tr>
<tr>
<td>Format (0 = PDA, 1 = paper)</td>
<td>23.31***</td>
</tr>
<tr>
<td>Format × item order</td>
<td>0.08***</td>
</tr>
<tr>
<td>Format × sex item</td>
<td>26.93***</td>
</tr>
<tr>
<td>Format × substance use item</td>
<td>–1.32</td>
</tr>
<tr>
<td>$R^2$</td>
<td>.91***</td>
</tr>
</tbody>
</table>

Note: PDA = personal digital assistant.

***p < .001.
In terms of the types of items in the assessment, after accounting for the other factors under investigation, items pertaining to sexual behavior were more likely to be missing than other types of items, whereas items about substance use may be less likely to be missing. Our initial models showed that these effects were similar for the PDA and paper assessments, with about a 1% change in missingness depending on item content. However, follow-up models that used alternative strategies for selecting items and calculating their percentage missingness showed that the significant positive association between sexual content and missingness was especially strong for participants in the paper-based assessment. Our estimates were that an item related to sexual behavior in a paper-based assessment would be missing between 17% and 27% more responses than the same type of item in a PDA assessment. The variability in this finding across alternative ways of structuring the data underscores the complexity of making direct comparisons between different types of assessments. However, it also suggests that PDAs may be especially helpful for data collection in the area of adolescent sexuality.

We believe that the PDA effects on missingness that we have demonstrated are of practical significance to evaluation researchers. The effects of assessment format and its interactions with item order and sexual content have the potential to alter rates of item missingness by anywhere from 10% to 30% or more, depending on the total number of items and the order in which content appears. In analyses that use only cases with complete data, this level of missing data reduces power and potentially yields biased parameter estimates. Although there are missing data procedures that can help correct for these issues, they do increase the effort and expertise required for data analysis. In contrast, data that are missing less than 5% of responses, as we had in our PDA assessment, can be submitted to complete case analysis with negligible detriment to inference (Graham et al. 2003). Therefore, the move from paper assessments to PDAs may translate into a savings of human and financial resources at the analysis stage of evaluation research.

Limitations and Future Directions

First, it is important to note that our two samples were not randomly drawn from the same population. Although the two samples were students of the same age from the same schools, there were differences between them in terms of cohort, race, and language spoken. Also, the items administered to each sample were not identical. In particular, the paper assessment was given as part of a pilot study; it is possible that missingness was reduced in the subsequent PDA assessment because of fine tuning to data collection procedures. In short, not all extraneous factors were controlled in
this design. To the extent that these factors are related to item missingness, our results could be biased.

However, we would argue that the samples are similar enough that valid conclusions can still be drawn from their comparison. It is unlikely that a 1-year cohort effect could account for the sizable differences between assessment formats. Also, although there may be some racial and language differences in the sample, these appear to have relatively small implications for missingness. By exponentiating the coefficients in Table 2, we see that Black students have, on average, three or more missing responses in their surveys than colored students. During the course of a 200-item survey, this works out to be less than 2% of possible responses. Students who take their survey in Afrikaans are missing an average of two or more responses than those who take their surveys in English. These person-level effects are much smaller than the effects of item survey format and most item-level characteristics demonstrated in this study.

Another potential confound is that participants in PDA assessment were more familiar with the survey instrument. These students, who were part of a full research trial with a pretest/posttest design, had already taken an identical survey approximately 6 months previously. However, the design of the paper-based pilot study was posttest only, so these students were seeing the survey for the first time. It is possible that students who completed paper assessments had higher rates of missingness than those who completed the PDA assessments because they were not completely familiar with the survey procedures. However, we believed it was important to reduce developmental differences between the two samples by using assessments from the same point in time (end of eighth grade). In addition, the rates of missingness in the PDA pretest, which represents those participants’ first exposure to the survey instrument, were still very low (average of 1.1% for each item). This implies that survey familiarity cannot account for much of the difference in missingness between paper and PDAs that was found in this study.

One of our possible explanations for differential missingness was that PDAs are more novel (and hence more motivating to use) than paper assessments. This raises the concern that as participants become more familiar with using PDAs, they may be decreasingly motivated to provide complete data. However, our data show that this is not the case. We have collected six waves of PDA data in our full intervention trial, and average missingness at Time 6 (end of 10th grade) is less than 0.4%. This suggests that PDAs can be effective data collection tools even in longitudinal research.

Although PDAs were found to be superior for eliciting nonmissing responses, we have not formally examined the relative quality of data that
were collected. It is possible that some participants are just keying in “something” (other than “refuse to answer”) as a way to get through a lengthy survey quickly while still giving the appearance of having provided useable data. However, we are encouraged by previous research showing that PDA use does not negatively affect data quality (e.g., Bernhardt et al. 2005; Fletcher et al. 2003; Forster et al. 1991; Green et al. 2006; LaBrie et al. 2006; Tapp et al. 2006; Treadwell 2006; Zwarenstein et al. 2006). Also, the variability in our PDA data may speak to their quality. Providing the same response (e.g., Option 1) to each survey item would be the simplest way of “rushing through” an assessment while still providing data. However, an examination of the last 15 items in our PDA survey showed that no participants entered the same response option for all of these questions. Therefore, we have no reason to believe that our PDA data are of a quality inferior to paper, although we acknowledge that this is an issue that can be explored in greater depth.

These limitations aside, we have strong evidence for the assertion that PDAs serve to reduce missingness during the course of long survey assessments. This underscores the utility of PDAs for evaluation research with adolescents.

Notes

1. There is one type of paper-based assessment for which the preceding discussion of relative advantages and disadvantages does not apply. Optical mark recognition forms (e.g., Scantrons) require a participant to make pencil marks on a form to correspond with their response choices. These marks are then read by an optical scanner, and the corresponding data can be made available electronically. Although this assessment mode is still unable to accommodate multiple languages or skip patterns in the ways that PDAs can, optical mark recognition data can be available for analysis more quickly than other forms of paper-based assessment.

2. Strictly speaking, a percentage outcome violates certain assumptions of linear regression in that it is skewed toward 0 and bounded by 0 and 100. Therefore, Model 4 was recomputed as a binomial logistic regression predicting the number of missing responses. The resultant coefficients were in the same direction as those appearing in Table 4, and all retained their statistically significant or nonsignificant status.

References


Lori-Ann Palen is a doctoral candidate in human development and family studies at the Pennsylvania State University. She is also a predoctoral fellow in the Prevention and Methodology Training Program, funded by the National Institute on Drug Abuse (NIDA). She currently serves as a project coordinator for HealthWise South Africa, where her research focuses on risk behavior prevention, research methodology, and adolescent free-time activities.

John W. Graham is a professor of biobehavioral health at the Pennsylvania State University. His research relates mainly to evaluation of health promotion (physical activity) and disease prevention programs. His recent work focuses on missing data analysis and design and applications of structural equation modeling.

Edward A. Smith is the director of evaluation research for the Prevention Research Center at the Pennsylvania State University. He also serves as the program director for the Center’s
NIDA-funded Prevention and Methodology Training Program. He is the co-principal investigator of the HealthWise South Africa Project.

**Linda L. Caldwell** is a professor of recreation, park, and tourism management and human development and family studies at the Pennsylvania State University. She is the co-principal investigator of the HealthWise South Africa Project. Most of her research centers on the intersection of youth development, leisure and health, and prevention. She primarily focuses on interventions that develop youth competencies, promote healthy lifestyles, and reduce risky behavior in and through leisure.

**Catherine Mathews** is a senior scientist for Health Systems Research Unit of the South African Medical Research Council. She is also an honorary lecturer in the Department of Public Health and Primary Health Care at the University of Cape Town.

**Alan J. Flisher** is a professor of psychiatry and mental health at the University of Cape Town, head of the Division of Child and Adolescent Psychiatry at University of Cape Town and the Red Cross War Memorial Children’s Hospital, director of the Adolescent Health Research Institute at University of Cape Town, and adjunct professor at the Research Centre for Health Promotion, University of Bergen, Norway. His main research interests are in adolescent mental health and risk behavior and mental health services research.