Effectiveness of Prompts on Fourth-Grade Children’s Dietary Recall Accuracy Depends on Retention Interval and Varies by Gender

Suzanne D Baxter,1,9* Albert F Smith,7,9 David B Hitchcock,6 Caroline H Guinn,5 Julie A Royer,5 Kathleen L Collins,5 Alyssa L Smith,5 Megan P Puryear,5 Kate K Vaadi,5 Christopher J Finney,5 and Patricia H Miller8

Abstract

Background: Dietary recall accuracy is related to retention interval (RI) (i.e., time between to-be-reported meals and the interview), and possibly to prompts. To the best of our knowledge, no study has evaluated their combined effect.

Objective: The combined influence of RI and prompts on children’s recall accuracy was investigated in this study. Two RIs (short [prior-24-h recall obtained in afternoon] and long [previous-day recall obtained in morning]) were crossed with 4 prompts [forward (distant-to-recent), meal-name (breakfast, lunch, etc.), open (no instructions), and reverse (recent-to-distant)], creating 8 conditions.

Methods: Fourth-grade children (n = 480; 50% girls) were randomly selected from consenting children at 10 schools in 4 districts in a southern state during 3 school years (2011–2012, 2012–2013, and 2013–2014). Each child was observed eating school-provided breakfast and lunch, and interviewed one time under 1 of the 8 conditions. Condition assignment was constrained so that each had 60 children (30 girls). Accuracy measures were food-item omission and intrusion rates, and energy correspondence rate and inflation ratio. For each measure, linear models determined effects of RI, prompt, gender, and interactions (2-way, 3-way); race/ethnicity, school year, and district were control variables.

Results: RI (P values < 0.015) and prompt (P values < 0.005) were significant for all 4 accuracy measures. RI × prompt (P values < 0.001) was significant for 3 accuracy measures (not intrusion rate). Prompt × gender (P = 0.005) was significant for omission rate. RI × prompt × gender was significant for inclusion rate and inflation ratio (P values < 0.001). For the short vs. long RI across prompts and genders, accuracy was better by 33–50% for each accuracy measure.

Conclusions: To obtain the most accurate recalls possible from children, studies should be designed to use a short rather than long RI. Prompts affect children’s recall accuracy, although the effectiveness of different prompts depends on RI and varies by gender: at a short RI, the choice of prompts has little systematic effect on accuracy; whereas at a long RI, reverse prompts may elicit the most accurate recalls. J Nutr 2015;145:2185–92.

Keywords: 24-h recall, children, retention interval, prompts, gender

Introduction

Data on children’s dietary intake, including intake at school, are obtained in both national surveys and research studies (1, 2). Although parents may report their elementary school children’s intake, the accuracy of such reports is questionable because parents lack first-hand knowledge of their children’s intake at school (3–5). Children in upper elementary grades have provided 24-h recalls without parental help for various research projects (2, 6–10), but validation studies have identified errors—both omissions and intrusions—in children’s recalls (5, 11–23).

To improve children’s dietary recall accuracy, factors that influence accuracy must be investigated. One certainly influential factor is retention interval (RI) (10) (time between meals to be reported and the interview): validation studies show that children’s
recall accuracy is better when the RI is shorter (12, 14, 21–25). The RI for a 24-h recall depends on choices of both the target period (e.g., prior 24 h [24 h immediately preceding the interview] and previous day [midnight to midnight of the day before the interview]) and interview time of day (e.g., morning, afternoon, and evening). Prompts used during the first pass of a 24-h recall may influence accuracy, too: individuals instructed to recall intake for a 24-h period initially may be prompted to do so in forward order (from the beginning of the target period to the end), reverse order (from the end to the beginning), or by meal name (e.g., breakfast and lunch), or may be given no specific instructions (i.e., open prompts). Few validation studies, to our knowledge, have investigated the effect of prompts on children’s recall accuracy (17, 26), and we know of only one small unpublished study, which we conducted, of the combined effects of RI and prompts on recall accuracy: in that study, 24 fourth-grade children were observed eating school breakfast and lunch, and then interviewed under 1 of 4 conditions created by crossing 2 RIs [short (prior-24-h-afternoon) and long (previous-day-morning)] with 2 prompts (forward and reverse), with 3 boys and 3 girls per condition. Results showed an RI × prompt interaction on 2 measures: recalls were most accurate for the short RI with reverse prompts, and for the long RI with forward prompts. Despite the small sample, results showed an RI × prompt × gender interaction on one measure: girls were most accurate for the long RI with either reverse or forward prompts, whereas boys were most accurate for the short RI with reverse prompts, but for the long RI with forward prompts.

Any 24-h recall has an RI, and all 24-h recall protocols use prompts. Both RI and prompts are choices made by researchers and/or practitioners, so understanding their separate and combined effects should inform the design of methods or studies to improve recall accuracy.

This article describes a cross-sectional validation study that investigated the effects of prompts on fourth-grade children’s dietary recall accuracy, and whether and how these effects depend on RI. Recall accuracy for 2 consecutive school meals was examined under 8 conditions created by crossing 2 RIs [short (prior-24-h recall in the afternoon) and long (previous-day recall in the morning)] with 4 prompts [forward (distant-to-recent), meal-name (breakfast, etc.), open (no instructions), and reverse (recent-to-distant)] (Supplemental Figure 1). As discussed in the Methods, 3 of these prompts are used in prominent 24-h recall protocols. Reported school-meal intake was validated with direct meal observation of school breakfast and lunch. Accuracy was assessed for food items and for energy.

There were 4 hypotheses, with the first 2 based on our small unpublished validation study: 1) the joint effect of RI and prompts on accuracy would overshadow the effect of either by itself; 2) the joint effect of RI and prompts would differ for boys and girls (i.e., an RI × prompt × gender interaction); 3) accuracy would be better for the short rather than long RI (12, 14, 21–25); and 4) interview length would be greater for the short rather than long RI (27–29), and when open prompts were used instead of the other 3 prompts.

Methods

The University of South Carolina’s Institutional Review Board approved the study. Written parental consent and child assent were obtained.

Selection of 8 interview conditions

The three prominent 24-h recall multiple-pass protocols are the Nutrition Data System for Research (NDSR) protocol (30), the Automated Multiple-Pass Method (AMPM) (31), and the Automated Self-Administered 24-Hour recall (ASA24) (32). Each protocol was designed to target the previous day. However, since 2013, the NDSR has allowed researchers to target the prior 24 h (although prompts embedded in NDSR software refer to the previous-day target period), and since February 2014, the ASA24 has included an option to target the prior 24 h [because of our research (AF Subar, National Cancer Institute, personal communication, 2014)]. The 3 protocols use different prompts during the first pass to obtain a list of foods eaten—the NDSR uses forward prompts, the AMPM uses open prompts by asking respondents to report intake freely without providing specific prompts, and the ASA24 uses meal-name prompts.

Another publication (28) details the rationale for our 8 conditions. For the 2 RIs, we used the results of our 2009 study of children’s recall accuracy for school meals (21) to pair each of 2 target periods (the less often-used prior-24-h and the more commonly used previous-day) with the interview time for which accuracy was best among examined interview times; this was afternoon for the prior-24-h target period and morning for the previous-day target period. In this article, these are the “short” and “long” RIs, respectively. We chose to investigate the prompts used in the 3 prominent protocols described earlier, and reverse prompts, which have been examined in dietary (17) and other survey methodology studies (33–35).

Sample size

To estimate required sample sizes, we used omission and intrusion rates from school-meal reports in fourth-grade children’s 24-h recalls in past studies (12, 17, 21). Assuming α = 0.025, we determined that a total of 480 children, with 30 children in each of 16 cells (2 RIs × 4 prompts × 2 genders), were required to detect main effects with 77–100% power, 2-way interactions with 99–100% power, and the 3-way interaction with 81% power. An α of 0.025 was assumed to protect against inflated Type I error in the unanticipated event that omission and intrusion rates were highly correlated. Power calculations were based on Neter et al. (36) and Winer (37).

Participants and design

Data were collected over 3 consecutive school years from consented children from all regular fourth-grade classes at 10 public elementary schools in 4 districts around Columbia, South Carolina. Data were collected during school year 2011–2012 in 6 schools, school year 2012–2013 in the same 6 schools plus 4 additional schools, and school year 2013–2014 in 4 schools involved during both previous school years and 1 school involved only during the previous school year. In 2 districts, these schools were the only elementary schools; in the other 2 districts, schools were identified based on high student participation in school-meal programs and the district’s permission for outside research studies. In all 4 districts, for each school year, permission to collect data was obtained from the district and school principals.

At the start of each school year, researchers visited each fourth-grade class at each school and invited children to participate by distributing assent and consent forms, reading the assent form aloud, asking questions to confirm children’s understanding of it, and responding to children’s questions. Researchers returned to each class 2–3 d later, at which time children who returned forms signed by their parents received small prizes irrespective of participation decisions. For the 3 respective school years, of the 569, 782, and 429 children invited to participate, 407 (71.5%), 521 (66.6%), and 280 (65.3%) agreed. Distributions of race/ethnicity and gender of the children who agreed to participate (range by school year: 47–54% African American, 35–42% White, 6–8% Hispanic, and 3–5% other; 47–55% girls) were similar to those of the children invited (46–50%, 36–43%, 6–8%, and 4–8%; 45–52%, respectively). Parents reported children’s race/ethnicity and gender to the schools, and the schools provided information to the researchers.

For each school year, consented children in the highest 2.5% and lowest 2.5% of the age distribution, who could have repeated or skipped a grade, were neither observed nor interviewed. From the consented children within the center 95th percent of the age distribution, children were randomly selected with the constraint that one-half were girls. Each sampled child was observed eating 2 consecutive school-provided meals—breakfast and lunch—and interviewed to obtain a 24-h recall under 1 of the 8 conditions yielded by crossing 2 RIs with 4 prompts
Assignment to condition was random with the constraint that data collection continued until each condition had 60 children (one-half girls). Insofar as possible within school year and school, random assignment to condition (by gender) was balanced across classes.

Neither children nor school staff knew in advance when observations and/or interviews would occur, nor to which interview condition any selected child would be assigned. Because more children were recruited than needed, children could determine neither who specifically was being observed at any meal nor who would be interviewed later. Although any child was interviewed only one time, children, when recruited, were told that they might be interviewed 0–2 times, so being interviewed did not indicate to a child that he or she would not be interviewed again.

Direct meal observations
Three researchers observed school-provided meals according to an established protocol with training through practice (12, 15–17, 21–23). Before data collection and weekly throughout data collection each school year, interobserver reliability was assessed for each pair of observers according to established procedures (12, 16, 17, 21–23, 38).

During data collection for breakfast for the 3 school years, interobserver reliability was assessed on 44, 72, and 23 children: mean agreement between observers to within one-fourth of a serving on amounts eaten was 99%, 99%, and 100%. During data collection for lunch for the 3 school years, interobserver reliability was assessed on 44, 80, and 24 children: mean agreement between observers to within one-fourth of a serving on amounts eaten was 98%, 97%, and 100%. These levels of agreement are excellent (39, 40).

During each meal observation, a researcher observed 1–3 children simultaneously during a regular meal period. Children were seated in their school’s typical arrangement. Observations covered entire meal periods to account for food trading (18, 41). Researchers used paper forms to record items and amounts eaten in servings of standard school-meal portions.

Dietary recall interviews
Four researchers conducted private interviews at schools during school year 2011–2012; 3 of these researchers conducted interviews during school years 2012–2013 and 2013–2014. Interviews for the short and long RIs were conducted Tuesdays through Fridays after lunch and breakfast, respectively. Three of the 4 interviewers also conducted observations, but a child’s interviewer had never observed the meals about which that child was interviewed. Interviewers were trained with the use of modeling, practice, and quality control assessment of practice interviews. Supplemental Figure 2 details each multiple-pass protocol. In the prior-24-h protocol passes for all 4 prompts, children were to report intake first for the interview day, and then for the previous day from 24 h before the interview’s start time; this approach, described by Buzzard (42), has been used with children (12, 21–23, 25). Children used qualitative terms to report amounts eaten in servings of standard school-meal portions (12, 15–17, 21–23). Interviewers used paper forms to document beginning and ending times, and information reported by children. Interviews were audio-recorded and transcribed. Each interview child was mailed a $10 check.

Quality control was applied to each interview according to established procedures (12, 16, 17, 21–23, 43) by having a noninterviewing researcher who had never observed the meals present at a child’s meal to observe the interview (42, 43). This interviewer had never observed the meals on which the interviewing child’s meal portions were typical of the child’s school’s typical arrangement. Observations covered entire meal periods. The protocol was recorded on a 60-min video, and the video became the basis for transcript creation, which was based on an established protocol (12, 16, 17, 21–23). Interviewers used paper forms to record items and amounts eaten in servings of standard school-meal portions.

Outcome measures
Seven response measures were examined—3 were characteristics of the interview and what was reported, but not the accuracy of what was reported, and 4 were accuracy measures. The 3 nonaccuracy measures were weighted number of items observed eaten, kilocalories observed eaten, and interview length (Supplemental Table 1). The 4 accuracy measures were food-item omission rate and intrusion rate, and energy correspondence rate and inflation ratio (Table 1).

Nonaccuracy measures. A meal-component weight was assigned to each item observed eaten at a school meal, regardless of quantity, with combination entrée (e.g., spaghetti with meat sauce) = 2, condiment (e.g., ketchup) = 0.33, and other meal components = 1 (12, 15–17, 21–23). These weights were summed over each child’s 2 school meals to obtain weighted number of items observed eaten.

For each item observed eaten, the amount observed eaten was quantified as servings of standard school-meal portions with none = 0, taste = 0.10, little bit = 0.25, half = 0.50, most = 0.75, and all = 1, and the actual number of servings if >1 (12, 15–17, 21–23). Obtaining kilocalories for a standard school-meal portion from the NDSR database, energy (in kilocalories) observed eaten was summed over each child’s 2 school meals.

Each 24-h recall interview’s length in minutes was calculated by subtracting beginning time from ending time. Interview length is useful for such practical purposes as scheduling research staff time and estimating class time that children will miss for interviews (6, 23).

Accuracy measures. Although children reported intake for a 24-h period, accuracy was assessed for only the school-meal parts of recalls because only school meals were observed. For meals in children’s 24-h recalls to be treated as reports about school meals, children had to identify “school” as the location at which items were eaten, refer to breakfast as “school breakfast” or “breakfast,” refer to lunch as “school lunch” or “lunch,” and report meal times to within 1 h of observed meal times (12, 15–17, 21–23). Any child who reported no meals that satisfied criteria to be considered school meals was excluded from analyses for the accuracy measures.

To quantify item-reporting accuracy, for each meal for each child, the set of food items reported eaten was compared with the set of food items observed eaten. According to an established classification system (11–13, 15–17, 21–23, 44–47), foods in both sets were matches, foods reported but not observed were intrusions, and foods observed but not reported were omissions.

To quantify item-reporting accuracy, for each meal for each child, the set of food items reported eaten was compared with the set of food items observed eaten. According to an established classification system (11–13, 15–17, 21–23, 44–47), foods in both sets were matches, foods reported but not observed were intrusions, and foods observed but not reported were omissions.

As in past studies (11–13, 15–17, 21–23, 44–46), we assigned a weighting to each match, omission, and intrusion according to the weighting system described earlier; summed weighted matches, omissions, and intrusions over each child’s 2 school meals; and calculated the child’s food-item omission rate—the percentage of weighted items observed eaten at school meals but unreported (Table 1, footnote 4)—and food-item intrusion rate—the percentage of weighted items reported eaten at school meals but unobserved (Table 1, footnote 6). Omission rates and intrusion rates may range from 0–100%; smaller values indicate better accuracy.

To report amounts eaten, children used qualitative terms for servings of standard school-meal portions—none, taste, little bit, half, most, all, and the actual number of servings if >1; for analyses, these were coded as 0, 0.10, 0.25, 0.50, 0.75, 1, and the actual number of servings if >1, respectively (12, 15–17, 21–23). For each reported item, the NDSR database was used to obtain kilocalories for a standard school-meal portion. Children’s accuracy for reporting energy (kilocalorie) intake was assessed as in past studies (23, 25, 48–51). For each match, either 1) the reported amount corresponded exactly to the observed amount, 2) the reported amount corresponded to part of the observed amount and the rest of the observed amount was unreported, or 3) part of the reported amount corresponded to the observed amount and the rest was over-reported. For each omission, the entire observed amount was unreported. For each intrusion, the entire reported amount was over-reported. Each corresponding, unreported, and over-reported amount was multiplied by the per-serving kilocalorie value to obtain corresponding, unreported, and over-reported kilocalories for each item. These were summed over each child’s 2 school meals and used to calculate energy correspondence rate—the ratio of corresponding kilocalories to observed kilocalories, expressed as a percentage (Table 1, footnote 7), and energy inflation ratio—the ratio of over-reported kilocalories to observed kilocalories, expressed as a percentage (Table 1, footnote 8). Correspondence rate may range from 0–100%; greater values indicate
TABLE 1  Descriptive statistics for 4 accuracy measures by RI, gender, and prompt from a dietary recall validation study in fourth-grade children\(^1\)

<table>
<thead>
<tr>
<th></th>
<th>Forward prompts</th>
<th>Meal-name prompts</th>
<th>Open prompts</th>
<th>Reverse prompts</th>
<th>Overall</th>
<th>RI</th>
<th>Prompt</th>
<th>Gender</th>
<th>Prompt × Gender</th>
<th>3-way</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Children in analyses for accuracy measures, n</strong></td>
<td>114 118 113 110 455</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Short RI, n</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total, short RI</td>
<td>59 60 57 57 233</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boys</td>
<td>30 30 29 28 117</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Girls</td>
<td>29 30 28 29 116</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Long RI, n</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total, long RI</td>
<td>55 58 56 53 222</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boys</td>
<td>26 29 27 24 106</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Girls</td>
<td>29 29 29 29 116</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Food-item omission rate(^2-4)</strong></th>
<th>Overall</th>
<th>Short RI</th>
<th>Boys</th>
<th>Girls</th>
<th>Long RI</th>
<th>Boys</th>
<th>Girls</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>43.6 ± 1.6(^5)</td>
<td>31.6 ± 1.2</td>
<td>18.8 ± 0.9</td>
<td>21.0 ± 2.6</td>
<td>16.6 ± 3.3</td>
<td>39.5 ± 3.7</td>
<td>54.0 ± 1.9</td>
</tr>
<tr>
<td><strong>Energy correspondence rate(^3,7)</strong></td>
<td>Overall</td>
<td>Short RI</td>
<td>Boys</td>
<td>Girls</td>
<td>Long RI</td>
<td>Boys</td>
<td>Girls</td>
</tr>
<tr>
<td></td>
<td>50.3 ± 2.7</td>
<td>60.5 ± 2.1</td>
<td>59.5 ± 3.7</td>
<td>61.5 ± 3.1</td>
<td>40.2 ± 3.6</td>
<td>40.7 ± 2.1</td>
<td>39.7 ± 7.3</td>
</tr>
</tbody>
</table>

| **Energy inflation ratio\(^7,8\)** | Overall | Short RI | Boys | Girls |
|                                    | 24.3 ± 2.9 | 16.2 ± 2.2 | 14.7 ± 1.6 | 17.8 ± 4.1 |

\(^1\) RI, retention interval.

\(^2\) Weights were assigned by meal component with combination entrée (e.g., spaghetti with meat sauce) = 2, condiment (e.g., ketchup) = 0.33, and remaining meal components = 1.

\(^3\) Amounts eaten were observed, reported, and scored in servings of standard school-meal portions as none = 0, taste = 0.10, little bit = 0.25, half = 0.50, most = 0.75, all = 1, and the actual number of servings if >1.

\(^4\) Calculated for a child as [sum of weighted omissions/(sum of weighted omissions + sum of weighted matches)] × 100. Values range from 0–100%; smaller values indicate better accuracy.

\(^5\) Least squares means ± SEMs (all such values).

\(^6\) Calculated for a child as [sum of weighted intrusions/(sum of weighted intrusions + sum of weighted matches)] × 100. Values range from 0–100%; smaller values indicate better accuracy.

\(^7\) Calculated for a child as (corresponding amounts in kilocalories from matches/observed amounts in kilocalories) × 100. Values range from 0–100%; greater values indicate better accuracy.

\(^8\) Calculated for a child as [over-reported amounts in kilocalories from matches and intrusions]/observed amounts in kilocalories] × 100. Values start with 0% but have no upper bound; smaller values indicate better accuracy.

Better accuracy. Inflation ratio would be 0% if no parts of matches were over-reported and there were no intrusions; inflation ratio has no upper bound (because there is no limit on what a child can report) and smaller values indicate better accuracy. Correspondence rate and inflation ratio sum to the frequently reported ratio of reported energy to observed energy (51).
Analyses
Generalized estimating equation methodology accounted for a possible correlation in response values within the same interviewer; an exchangeable correlation structure was assumed. For all models, residual analysis checked assumptions of constant error variance and normality of errors. Simultaneous Wald tests jointly tested for factor effects in each model. The Benjamini-Hochberg procedure (52) adjusted for simultaneous tests on multiple factors and in multiple models, controlling the false discovery rate at \( \alpha = 0.05 \). All P values reported are adjusted. Analyses used SAS/STAT (version 9.4, 2002–12). Various linear models were fit to determine main effects of RI, prompt, and gender, their 2-way interactions, and the 3-way interaction on 3 nonaccuracy measures and 4 accuracy measures. Race/ethnicity, school year, and district were control variables; for the nonaccuracy measures, observer was also a control variable. Inflation ratio was square-root transformed to help satisfy the normality assumption.

Results
Of 480 children (one-half girls, by design) with interviews included in analyses, there were 261 African American (54%), 150 White (31%), 47 Hispanic (10%), and 22 other (5%).

For accuracy measures, we predicted that, regardless of prompt, performance would be better when the RI was short rather than long. Of particular interest was whether prompts affected performance, and whether and how any effect of prompts differed at the 2 RIs. We made no predictions about the association between performance and gender, but powered the study to detect a gender effect and populated each condition with equal numbers by gender. Of interest was whether any joint effect of RI and prompt differed systematically by gender. The tables reflect these interests: the primary division for each measure in the tables is by RI; at each RI, results are shown for each of the 4 prompts for boys and for girls.

Recalls from 25 children (8 girls) failed to meet criteria for any reported meal to be classified as a school meal [7 short RI (1 forward, 3 open, and 3 reverse) and 18 long RI (5 forward, 2 meal-name, 4 open, and 7 reverse)]. Most of these children were assigned to the long RI, within which they were distributed approximately equally over the 4 prompts. Recalls from these 25 children were analyzed for the 3 nonaccuracy variables, but not for the 4 accuracy variables.

Nonaccuracy measures. Least squares means for weighted number of items observed eaten at 2 school meals by RI, prompt, and gender are shown in Supplemental Table 1. None of the main effects, nor any interaction, was significant (all \( P > 0.12 \)). Taking into account the variation associated with the control variables, the weighted number of items observed did not vary systematically over conditions, consistent with what would be expected given random assignment of children to the 8 conditions.

Least squares means for kilocalories observed eaten at 2 school meals by RI, prompt, and gender are shown in Supplemental Table 1. None of the main effects, nor any interaction, was significant (all \( P > 0.08 \)). Combined with the results for weighted number of items observed, the fact that the \( P \) values for the RI × prompt and RI × prompt × gender interactions were between 0.08 and 0.09 does not meaningfully challenge the notion that intake did not vary systematically over conditions.

Least squares means for interview length (in minutes) by RI, prompt, and gender are shown in Supplemental Table 1. The main effects of RI and prompt, and the RI × prompt, prompt × gender, and RI × prompt × gender interactions, were significant (all \( P < 0.001 \)). Neither the main effect of gender (\( P = 0.92 \)) nor the RI × gender interaction (\( P = 0.80 \)) was significant. Overall, interviews for the short RI were 3 min lengthier than those for the long RI; also, overall, interviews with meal prompts were fastest and those with open prompts were slowest. The detailed pattern was more complex, with a 3-way interaction. With the short RI, interview length varied greatly over prompts by gender: boys' means ranged from 19.4 min (meal-name) to 25.8 min (open), whereas girls' means ranged from 17.4 min (meal-name) to 25.3 min (open). In contrast, with the long RI, interview length varied less over prompts for either gender: boys' means ranged from 17.2–19.4 min, and girls' means ranged from 17.4–19.9 min.

Accuracy measures. Least squares means for omission rate by RI, prompt, and gender are shown in Table 1. Omission rate is the unreported proportion of to-be-reported items; higher values indicate greater error. The main effects of RI and prompt, and interactions of RI × prompt (all \( P < 0.001 \)) and prompt × gender (\( P < 0.005 \)) were significant. The main effect of gender (\( P = 0.80 \)), and the RI × gender (\( P = 0.29 \)) and RI × prompt × gender (\( P = 0.80 \)) interactions, were not significant. Overall, omission rate was much lower with the short (31.2%) than with the long (54.1%) RI. For the RI × prompt interaction, with the short RI, omission rate varied little, ranging from 27.8% with open and 28.9% with reverse to 36.4% with meal-name prompts; with the long RI, omission rate varied more, from 44.4% with reverse to 61.4% with open prompts. For the prompt × gender interaction, girls' omission rate varied little across prompts (39.7–44.4%), but boys' omission rate ranged from 33.6% with reverse to 47.0% with open and 48.7% with meal-name prompts (results not shown).

Least squares means for intrusion rate by RI, gender, and prompt are shown in Table 1. Intrusion rate is the proportion of reported items not observed; higher values indicate greater error. The main effects of RI (\( P < 0.001 \)) and prompt (\( P = 0.005 \)) were significant, as were the prompt × gender and RI × prompt × gender interactions (both \( P < 0.001 \)). The main effect of gender (\( P = 0.49 \)), and the RI × prompt (\( P = 0.80 \)) and RI × gender (\( P = 0.34 \)) interactions, were not significant. Overall, intrusion rate was much lower with the short (17.2%) than with the long (35.6%) RI. For the 3-way interaction, with the short RI, boys' intrusion rate was lower with reverse prompts (10.8%) than with the other prompts (18.1–21.0%), whereas girls' intrusion rate varied little over prompts (15.9–18.9%). With the long RI, boys' intrusion rate ranged from 21.4% with reverse prompts to 50.8% with meal-name prompts, but girls' intrusion rate ranged from 23.5% with meal-name prompts to 39.2% with reverse and 39.4% with forward prompts.

Least squares means for correspondence rate by RI, prompt, and gender are shown in Table 1. Correspondence rate indicates the extent to which kilocalories from foods reported eaten match kilocalories observed eaten; higher values indicate better accuracy. The main effects of RI and prompt were significant, as was the RI × prompt interaction (all \( P < 0.001 \)). The main effect of gender (\( P = 0.99 \)), and the prompt × gender (\( P = 0.09 \)), RI × gender (\( P = 0.46 \)), and RI × prompt × gender (\( P = 0.79 \)) interactions, were not significant. Overall, correspondence rate was higher with the short (62.2%) than with the long (41.5%) RI. For the RI × prompt interaction, when the RI was short, correspondence rate was highest with open prompts (67.3%) and lowest with meal-name prompts (56.9%), whereas when the RI was long, correspondence rate was highest with reverse prompts (50.8%) and lowest with open prompts (35.4%).

Least squares means for inflation ratio by RI, prompt, and gender are shown in Table 1. Inflation ratio indicates the extent
to which over-reported kilocalories augment correctly reported kilocalories; higher values indicate greater error. The main effects of RI \( (P = 0.015) \) and prompt \( (P < 0.001) \) were significant, as were the RI \( \times \) prompt, prompt \( \times \) gender, and RI \( \times \) prompt \( \times \) gender interactions \( (3 \ P \ values < 0.001) \). Neither the main effect of gender \( (P = 0.32) \), nor the RI \( \times \) gender interaction \( (P = 0.92) \), was significant. Overall, inflation ratio was lower with the short \( (20.1\%) \) than with the long \( (30.5\%) \) RI. For the 3-way interaction, with the short RI, girls' inflation ratio varied relatively little over prompts \( (15.8\%–20.0\%) \), whereas boys' inflation ratio ranged from \( 14.7\% \) with forward prompts to \( 31.2\% \) with reverse prompts. With the long RI, inflation ratio varied considerably over prompts for both genders (although more for boys than girls), and prompts with the lowest and highest inflation ratios differed by gender; meal-name prompts had the highest inflation ratio for boys \( (30.1\%) \), but lowest for girls \( (16.3\%) \).

### Discussion

We investigated fourth-grade children's accuracy for reporting school breakfast and school lunch—obtained during 24-h recalls under 8 conditions constructed by crossing 2 RIs (short and long) with 4 prompts (forward, meal-name, open, and reverse)—to determine whether prompts affected recall accuracy, and to determine whether and how the effects of prompts depended on RI.

Hypothesis 1 was supported—the combined influence of RI and prompts on accuracy overshadowed the influence of either by itself. When the RI was short, omission rate was approximately equally best with open and reverse prompts; intrusion rate was best with reverse prompts; correspondence rate was best with open prompts; and inflation ratio was best with forward prompts. When the RI was long, omission rate, intrusion rate, and correspondence rate were all best with reverse prompts, but inflation ratio was best with open prompts. Accuracy was better with the short than long RI; with the short RI, none of the 4 prompts outperformed the others, whereas with the long RI, reverse prompts generally elicited the most accurate reports.

Hypothesis 2 was confirmed—the joint effect of RI and prompts differed by gender. Gender was not a significant effect for any measure, but it interacted with RI and/or prompt for 3 accuracy measures (omission rate, intrusion rate, and inflation ratio) and interview length. For omission rate, boys' accuracy was best with reverse and worst with open and meal-name prompts, but girls' accuracy varied little across prompts. For intrusion rate, for the short RI, boys' accuracy was best with reverse prompts, but girls' accuracy varied little across prompts; for the long RI, boys' accuracy was also best with reverse prompts, but girls' accuracy was best with meal-name prompts. For inflation ratio, for the short RI, boys' accuracy was best with forward prompts, but girls' accuracy varied little across prompts; for the long RI, boys' accuracy was best with open prompts, but girls' accuracy was best with meal-name prompts.

Hypothesis 3 was confirmed—accuracy was better when the RI was short rather than long; this is consistent with past findings with children \( (12, 14, 21–25) \), and is the most salient feature of the results. This superiority was found across the 4 prompts and for both genders for all 4 accuracy measures, although the magnitude of this difference depended on other factors. For recalls obtained with the short RI compared with the long RI, omission rates were better by \( 33–50\% \), intrusion rates were better by \( 50\% \), correspondence rates were better by \( 33\% \), and inflation ratios were better by \( 33\% \). Minimizing the RI between the meals to be reported and the interview time is the best strategy to maximize recall accuracy of which we are aware.

Hypothesis 4 was confirmed—interview length was greater \( 1 \) when the RI was short rather than long, and \( 2 \) when open prompts were used rather than the other 3 prompts. Recalls that used the short RI were 3 min longer than recalls that used the long RI \( (21.6 \ vs. 18.6 \ min) \). This was expected \( (27–29) \) and may be due to the segmented nature of the prior-24-h protocol passes, which, regardless of prompts, involved asking children to report intake first for the interview day, and then for the previous day beginning \( 24 \) h before the interview's start time. Recalls that used open prompts were \( 2.5–4.6 \) min longer on average than recalls that used the other prompts. This may be due in part to the nature of the open prompt multiple-pass protocol, which requires respondents to identify, for each item reported, a name for the meal at which it was eaten, and time and location for that meal (Supplemental Figure 2). Interview length does not reflect accuracy, but it is a design consideration in studies that use dietary recalls, especially with children.

We found that accuracy was best for the short RI: averaging across prompts and genders, the omission rate was \( 31\% \) and the intrusion rate was \( 17\% \). Do these seemingly high levels of error imply that \( 24\)-h recalls from children are unusable? We cannot answer this question, but we can make 2 pertinent remarks. The first involves comparing children's accuracy to that of adults: in a recent validation study with adults \( (53) \), averaging across ASA24 and AMPM previous-day 24-h recalls, the omission rate was \( 19\% \) and intrusion rate \( (\text{estimated from published data}) \) was \( 8\% \). That study's to-be-reported meals were eaten at a special center at which participants served themselves from a buffet at 10-min intervals, so these rates plausibly estimate the upper bound of accuracy for adults; under more typical conditions, adults' accuracy would likely be worse. Second, development of alternative dietary intake methods is desirable, but because \( 24\)-h recalls are widely used, optimizing this method is also a priority.

Future study of prompts' effect on dietary recall accuracy might examine a multiple-pass protocol with mixed-order prompts. Eyewitness testimony research has shown that more information was retrieved by witnesses who recalled a crime scene scenario \( 2 \) times—once in reverse order and once in forward order—than by witnesses who recalled \( 2 \) times in forward order \( (54, 55) \). For example, a reverse-order pass could be inserted into the NDSR's forward-prompt protocol; where such a pass might be placed to optimize performance would require research.

One limitation of this study was that it was not logistically possible to assess the validity of reports of intake for an entire \( 24 \) h in a way that was economical, generalizable, and no more than minimally reactive. Another limitation was homogeneity in grade level, so that our understanding of the effects of prompts was limited to fourth-grade children. Eight conditions were required to test 2 RIs crossed with 4 prompts, and gender was a design variable; thus, adding even 1 additional grade to the design would have been crushingly costly. A third limitation is that accuracy results, as measured by correspondence rate, might be inflated by restricting observations to standard school-meal portions. However, this concern is not pertinent to results concerning omission rate and intrusion rate because they do not consider portion size: an intrusion is an intrusion regardless of the amount of the intruded item reported eaten \( (\text{e.g., 0.5 or 3 servings}) \), and an omission is an omission regardless of the amount of the omitted item that had been eaten \( (\text{e.g., 0.25 or 2 servings}) \).

Strengths include use of the “gold standard”—direct meal observation—in a typical setting to validate parts of \( 24\)-h recalls \( (56) \). We have shown with fourth-grade children that conclusions
about dietary recalls validated by school-meal observations may be
generalized to recalls by comparable children who are not observed (27, 29).
We recruited more children than needed for data collection, so children could not anticipate whether they would be interviewed. Quality control for observations and interviews was assessed regularly during data collection.

Twenty-four-hour recalls will undoubtedly continue to be used with children, and investigators will continue to strive to obtain the most accurate recalls possible. The most effective strategy currently known for achieving this—with children—is to design studies that use a shorter rather than a longer RI. Prompts affect recall accuracy, although we found that the effectiveness of different prompts depends on RI and varies by gender: at the short RI, the particular choice of prompts has little systematic effect on accuracy, whereas at the long RI, reverse prompts, not used in any of the 3 prominent 24-h recall protocols (see Methods, first subsection), may elicit the most accurate recalls. In studies in which 24-h recalls are to be obtained only from girls or from boys, our findings for gender may influence choice of prompts.

Acknowledgments
SDB, AFS, and PHM designed the research; CHG, KLC, ALS, MPP, and KK conducted the research; DBH, JAR, and CJF analyzed the data; and SDB and AFS wrote the paper. All authors read and approved the final manuscript.

References


