Published in final edited form as:

Appl Cogn Psychol. 2008 December ; 22(8): 1038–1061. doi:10.1002/acp.1399.

Children's Dietary Recalls from Three Validation Studies: Types of Intrusion Vary with Retention Interval

Suzanne Domel Baxter^{1,*}, James W. Hardin¹, Albert F. Smith², Julie A. Royer¹, and Caroline H. Guinn¹

¹University of South Carolina, Columbia, SC, USA

²Cleveland State University, OH, USA

SUMMARY

Using previously collected data of fourth-grade children observed eating school meals and then interviewed, we categorized intrusions (food items reported but not observed eaten) as stretches (on the child's tray) or confabulations (not on the child's tray). We investigated intrusions, confabulations, and stretches, and the role of liking, at different retention intervals (morning interviews about the previous day's intake; evening interviews about that day's intake) and under different reporting-order prompts (forward; reverse). As retention interval between consumption and report increased, the likelihood 1) increased that reported items were intrusions, that reported items were confabulations, and that intrusions were confabulations; and 2) was constant that reported items were stretches. Results concerning reporting-order prompts were inconclusive. Liking ratings were higher for matches (reports of items observed eaten) than stretches, for confabulations than stretches, and for matches than omissions (unreported items observed eaten), but did not vary by retention interval or reporting-order prompts.

Keywords

children; dietary recalls; types of intrusion; retention interval; reporting-order prompts

Accurate assessment of dietary intake is challenging, especially among children (Baranowski & Simons-Morton, 1991; Frank, 1991). It is difficult for parents to accurately report their children's intake (Eck, Klesges, & Hanson, 1989), especially after children begin attending school (Emmons & Hayes, 1973). Thus, it is necessary to obtain dietary-intake information from children themselves.

A common method for collecting information about intake is the 24-hour dietary recall (24hDR), for which an interviewer asks a respondent about his or her intake for a 24-hour target period. Although data collection and processing are costly, 24hDRs do not require that respondents be literate, and the procedure is unlikely to alter intake (Buzzard, 1998).

Children have provided 24hDRs for national surveys (e.g., Gleason & Sutor, 2001; Moshfegh, Goldman, & Cleveland, 2005) and research studies (e.g., Luepker et al., 1996; Nicklas, Webber, Srinivasan, & Berenson, 1993; Weber et al., 2004). Children's 24hDRs have also been used to evaluate nutrition-education interventions (Contento, Randell, & Basch, 2002), and multiple

*Correspondence to: Suzanne Domel Baxter, PhD, RD, FADA, Institute for Families in Society, University of South Carolina, 1600 Hampton Street Suite 507, Columbia, SC 29208; Tel: (803) 777-1824 ext 12; Fax: (803) 777-1120, E-mail: sbaxter@gwm.sc.edu.

24hDRs by children have been used to assess the relative validity of food frequency questionnaires (e.g., Field et al., 1999; Jensen et al., 2004).

Validation of dietary recalls

In *validation* studies, dietary reporting accuracy is assessed by comparing *reported* information (e.g., 24hDRs) to *reference* information (e.g., meal observation) (Baxter, Smith, Hardin, & Nichols, 2007). (In *relative validation* studies, reference information consists of another self-report method such as food records.) To evaluate reporting performance at the food-item level, reported items may be classified as *matches*—items in the reference set—and *intrusions*—items not in the reference set. Reference items not matched by a reported item are *omissions* (Smith, 1991; Smith, Jobe, & Mingay, 1991).

Mertz (1992) asserted that *observation* is the best method for validating dietary recalls, and recommended that observations occur in a cafeteria-type setting familiar to subjects. Observations of school meals (breakfast, lunch) from school food service provide an excellent opportunity to obtain reference information to validate the school-meal parts of children's 24hDRs (Frank, 1991; Simons-Morton & Baranowski, 1991). In the United States, millions of children eat school meals each school day (U.S. Department of Agriculture, Economic Research Service, n.d.). Observations of children eating meals in private homes is intrusive (Caballero et al., 1998), obvious (Simons-Morton & Baranowski, 1991), unacceptable in some communities (Weber et al., 2004), and may cause substantial reactivity (Smith, 1999). Reactivity is less likely when observations are conducted at school where children are accustomed to being watched while eating (Baxter, Thompson, Litaker, et al., 2003; Simons-Morton & Baranowski, 1991) and where groups may be observed without revealing individuals being observed (Baxter, Thompson, Smith, et al., 2003). In most validation studies in which children provided oral 24hDRs without parental assistance, reference information was obtained by observing one or two school meals (Baranowski et al., 2002; Baxter et al., 2004; Baxter, Smith, Guinn, et al., 2003; Baxter, Thompson, Litaker, Frye, & Guinn, 2002; Baxter, Thompson, Litaker, et al., 2003; Baxter, Thompson, Smith, et al., 2003; Crawford, Obarzanek, Morrison, & Sabry, 1994; Lytle et al., 1993; Lytle, Murray, Perry, & Eldridge, 1998; Samuelson, 1970; Todd & Kretsch, 1986; Weber et al., 2004).

A salient characteristic of children's dietary recalls is the high percentage of reported items that are intrusions. A measure of this characteristic is the *intrusion rate*¹—the proportion of reported food items that was not eaten—with higher rates indicating worse reporting accuracy.

Types of intrusion

Memory researchers have classified reporting errors of commission, that is, intrusions or *false memories*, in various ways. For example, in analyzing children's reports about a magic show, Memon, Wark, Bull, and Koehnken (1997) distinguished between *errors*—incorrect descriptions of things that were present or events that did occur—and *confabulations*—reports of details that were not present or events that did not occur (see also Davis, McMahon, & Greenwood, 2005). In a comparable study with children, Jones and Pipe (2002) distinguished similarly between *distortions* and *intrusions*. In a study of free recall of lists of words with college students, Zaromb et al. (2006) distinguished between *prior-list intrusions*—incorrect words that had occurred on a previous list in the study—and *extra-list intrusions*—incorrect words that had not occurred on any list in the study. Each of these distinctions is between a

¹The intrusion rate is the complement of the output-bound performance measure of the accuracy-oriented correspondence approach to memory assessment (in contrast to the quantity-oriented storehouse approach, whose input-bound performance measure we call the match rate). See Koriat and Goldsmith (1994; 1996; see also Koriat, Goldsmith, & Pansky, 2000) for the distinction between these approaches to memory assessment, and see Smith et al. (1991, Appendix) for a discussion of the match and intrusion rates and their relationship.

reporting error that, in the judgment of the investigators, appears to be a misreport of something that actually occurred in the target context and a report of something that did not occur.

For this article, we distinguish between two types of intrusion in children's reports about school meals—stretches and confabulations. A *stretch* is an item that was on the child's tray for that meal, and a *confabulation* is an item that was not on the child's tray for that meal. Consider, for example, a child observed eating a specific school lunch, and his or her recall of that lunch. Suppose that 1) the items observed on the child's tray were chicken, broccoli, rice, applesauce, and white milk; 2) the child was observed eating chicken, rice, applesauce, and white milk; and 3) the child later reported having eaten some amount of chicken, broccoli, roll, and white milk for that lunch. The broccoli and roll are *intrusions*: The broccoli is a *stretch* because it was on the child's tray; the roll is a *confabulation* because it was not on the child's tray. (The chicken and white milk are *matches*; the rice and applesauce are *omissions*.)

Accounts of intrusions

Understanding the conditions and psychological processes that lead to reports of intrusions has become a major focus of research on memory (e.g., Brainerd & Reyna, 2005; Roediger, Balota, & Watson, 2001). In this article, we attempt to advance this understanding by examining the relationship between occurrence of types of intrusion in children's dietary recalls and (a) *retention interval*—the time between what is to be reported [e.g., a meal] and the report—which has consistently been demonstrated to affect intrusion rates, and (b) *reporting order*—the temporal order in which the events of a target period are to be reported—for which the impact on intrusion rates is less clear.

Mean intrusion rates for dietary reports have increased with retention interval in one relative validation study with adults and two validation studies with children. Smith et al. (1991) had adults keep food records for 2 or 4 weeks, and then report their intake for the recordkeeping period either at the end of that period or 2, 4, or 6 weeks later; reported items that were not recorded were classified as intrusions. Mean intrusion rates increased significantly with retention interval. Baxter, Thompson, Davis, and Johnson (1997) observed fourth-grade children eating school lunch and asked them to recall intake for that lunch within 1.5 hours of eating, the next morning, or on Monday morning (concerning Friday's lunch); mean intrusion rates increased significantly as the retention interval increased, with respective values of 5%, 13%, and 48%.² Baxter et al. (2004) observed fourth-grade children eating school breakfast and school lunch and then obtained a 24hDR about either the previous day (i.e., midnight to midnight of the day before the interview) or the prior 24 hours (i.e., from 24 hours before the interview to when the interview began), with the interview conducted either in the morning, afternoon, or evening. The mean intrusion rate was significantly greater for the school meal parts of 24hDRs about the previous day (54%) than the prior 24 hours (29%).³

Some evidence from research on eyewitness testimony (e.g., Geiselman & Callot, 1990) and autobiographical memory (e.g., Jobe et al., 1990; Loftus, Smith, Klinger, & Fiedler, 1992) indicates that the number of intrusions and mean intrusion rates may be smaller with reverse-than forward-order instructions. For dietary reports, evidence from one relative validation study with adults and one validation study with children is mixed concerning the effect of reporting-order prompts on intrusions. Armstrong et al. (2000) found that mean intrusion rates in the second of 2 back-to-back 24hDRs about a target day that was 4 days earlier did not depend on whether college students were instructed to recall in reverse or in forward order. All participants

²Children's baseline ability to recall items in a single lunch is available from the 89 children who provided 1.5 hour recalls (Baxter et al., 1997).

³Children's baseline ability to recall items eaten at school breakfast and school lunch during 24hDRs is available from the 10 children interviewed in the afternoon about the prior 24 hours (Baxter et al., 2004).

were instructed to report the first of the 24hDRs in forward order. Baxter, Thompson, Smith, et al. (2003) observed fourth-grade children eating 2 school meals (breakfast, lunch) on each of 2 days separated by ≥ 29 days, and then interviewed children in the morning on the day after each observation day about the previous day's intake. For each child, one interview used forward-order prompts and the other used reverse-order prompts. Significantly more boys had an intrusion rate $\leq 30\%$ with reverse-order prompts (54%) than forward-order prompts (40%).

Smith et al. (1991) suggested that their finding that mean intrusion rates increase with retention interval was consistent with the idea that, over time, an individual's general personal dietary knowledge is increasingly the basis of his or her reports. Elaborating on this, Smith and Jobe (1994) proposed that a person reporting dietary intake for a specific period samples candidate items from a categorically-organized memory and, with some probability, reports those items. Smith and Jobe assumed that memory contains permanent elements that reflect long-term dietary base rates, and episodic elements that are labeled during particular eating occasions, and that contingent on being sampled, the probability of reporting permanent elements is constant (but low) and the probability of reporting a labeled episodic element decreases over time since its labeling. These assumptions are compatible with Raaijmakers and Shiffrin's theory of probabilistic search of associative memory (1980, 1981). Simulations by Smith and Jobe (1994) were consistent with the empirical results.

Smith and Jobe (1994) did not distinguish between types of intrusion. However, if, during eating occasions, episodic elements are labeled for availability, and for having been eaten, then this model would predict that the probability of reporting a stretch, given items that could be reported as stretches, would decrease over time. The contribution of such items to overall reporting would depend on additional assumptions about the reporting process. Smith and Jobe did not model the number of items reported nor attempt to explain what determines how many items an individual reports in a dietary recall. However, for any number of items that an individual might report, the model predicts that, over time, the probability that an intrusion is a confabulation will increase, and that if individuals do not reduce the number of items they report as fast as episodic element labels decay, intrusion rate will increase, due primarily to an increase in the reported number of items that are confabulations. It is plausible that, in reporting intake for a meal or for 24 hours, individuals report some quota of items and so do not reduce the number of items they report as fast as episodic element labels decay. This model does not contain mechanisms that provide a basis for predictions about the effect of reporting-order prompts.

According to fuzzy trace theory (Brainerd & Reyna, 2004; Reyna & Brainerd, 1995), an experienced event is represented in memory as both a *verbatim trace*—with details specific to the event—and a *gist trace*—with interpretational information about the event; these traces are stored in parallel. Brainerd and Reyna proposed that verbatim traces are lost more rapidly than gist traces. Thus, reports about an experienced event obtained soon after the event may be based on retrieval of verbatim traces, but as the retention interval increases, individuals may rely increasingly on gist traces. For example, Jones and Pipe (2002) found, in a study of children's reports about a visit with a pirate, that the total number of reporting errors did not vary over retention intervals that ranged from immediate to six months, but that the composition of the errors varied: As retention interval increased, distortions decreased and intrusions increased. Jones and Pipe characterized these results as being compatible with the predictions of fuzzy trace theory; they supposed that the distortions were based on verbatim traces, and so decreased over time, and that intrusions were based on gist traces.

Applying the fuzzy trace framework to dietary recall, one would expect that an erroneous report of a verbatim trace of an item available, but uneaten, at a meal, would be more likely to occur at a short than at a long retention interval; in other words, stretch reporting should decrease as

retention interval increases. At the same time, one would expect confabulations to increase: If the gist trace that is retrieved is some type of schema of a meal, and if, as retention interval increases, subjects increasingly rely on gist traces for reporting, it is plausible that items necessary to complete the schema would be reported. To the extent that reverse-order prompts suppress the use of schemas in recall, and to the extent that forward-order schemas play a role in dietary recall, fuzzy trace theory might predict fewer intrusions with reverse-order prompts.

From the perspective of source monitoring (e.g., Mitchell & Johnson, 2000), items on a child's tray for a school meal may be construed as two sources—items eaten and items not eaten. Subsequent reporting of items eaten requires discriminating these sources. As retention interval increases, the difficulty of this discrimination may increase. A third source of information is items that have ever been eaten, particularly in the recent past. Failure to filter out this source would result in the reporting of items that are confabulations.

Finally, liking a food may facilitate its recall (Fries, Green, & Bowen, 1995). Given that stretches are items that were on the tray but were not eaten, it is reasonable to infer that the child likes those items less than items that were on the tray and were eaten. If a child reports items not eaten to fill a quota, or to complete a schema, it is reasonable to expect reporting of liked items. Finally, it is plausible that items eaten but not reported are liked less than items eaten and reported. Thus, we predict children's liking to be higher 1) for matches than for stretches, 2) for confabulations than for stretches, and 3) for matches than for omissions.

Synopsis of studies analyzed for this article

Using data from three validation studies in which fourth-grade children were observed eating school breakfast and school lunch and then interviewed to obtain a dietary recall, we conducted analyses to better understand intrusions in children's dietary recalls by investigating the occurrence of all intrusions and of two types of intrusions, as well as the role of liking, at different retention intervals and under different reporting-order prompts.

Study 1 addressed the consistency of children's dietary recall accuracy; each child was interviewed 1, 2, or 3 times (with ≥ 28 days between any two interviews) using forward-order prompts in the morning on the day after the observation day about the previous day's intake (Baxter et al., 2002). Study 2 addressed the effect of reporting-order prompts on children's dietary recall accuracy; each child was interviewed twice—once per reporting-order prompt condition (forward; reverse)—with ≥ 29 days between interviews, in the morning on the day after the observation day about the previous day's intake (Baxter, Thompson, Smith, et al., 2003). Study 3 addressed the effect of interview modality on children's dietary recall accuracy; each child was interviewed once, either in person or by telephone, with forward-order prompts in the evening of the observation day about that same day's intake (Baxter, Thompson, Litaker, et al., 2003). For this article, the two retention intervals were for interviews conducted approximately 21 hours (for Studies 1 and 2) and 7 hours (for Study 3) after the midpoint of the target period, and the two reporting-order prompt conditions were forward (in Studies 1, 2, and 3) and reverse (in Study 2). Although results of children's reporting accuracy from each study have been published, we have not previously investigated types of intrusion or liking ratings.

METHOD

Sample and general design

For each study and the current analyses, approval was obtained from institutional review boards for research involving human subjects, and child assent and parental consent to participate were obtained in writing prior to data collection. The Appendix summarizes details of the

studies that are relevant to the analyses reported in this article. The three studies were conducted during consecutive school years; each study was conducted during a single school year. No individual child was observed and interviewed for more than one of the three studies. For each study, children were recruited from all fourth-grade classes in 6 to 11 of 12 public elementary schools in one school district in a southern state. Schools were selected to permit a final sample for each study with equal numbers of Black and White girls and boys with high participation in school meals. Although the sex/race groups were represented approximately equally in each study, no study was powered to detect sex or race differences.

In schools where data were collected, most foods were served to children; *offer-versus-serve*, in which children may refuse a certain number of meal items, was not implemented.⁴ There were exceptions: Children had to take milk, but could select the fat content and flavor. Children could select limited quantities of condiments (e.g., mustard) when available (in single-serve packets). For breakfast each day, children chose either a cold breakfast (i.e., milk, juice, cereal) or hot breakfast (i.e., milk, juice or fruit, hot item [e.g., pancakes]). During lunch on most days, children could buy ice cream as an a la carte item.

Observations

Only children who obtained meals at school were observed because it can be difficult to unobtrusively identify meals brought from home (Simons-Morton et al., 1992). Entire school meal periods were observed due to trading of food items (Baxter, Thompson, & Davis, 2001; Simons-Morton & Baranowski, 1991). For each observed meal in Studies 1, 2, and 3, one of three, four, or three dietitians, respectively, observed one to three randomly-selected children simultaneously and recorded, for each child, items and amounts eaten in servings of standardized school portions. Children could see that an observer was present, but children did not know who, specifically, was being observed or would be interviewed. For each study, practice observations were conducted prior to data collection to familiarize children with an observer's presence (Simons-Morton & Baranowski, 1991). Training of observers for each study included modeling, practice, and assessment of inter-observer reliability (Baglio et al., 2004). Throughout data collection for each study, inter-observer reliability was assessed regularly and met established criteria (see the Appendix; Baranowski et al., 1986; Simons-Morton et al., 1992).

Interviews

For Studies 1 and 2, one of three or four dietitians, respectively, interviewed each child individually in person in the morning after breakfast on the day after school meals were observed; the target period for these interviews was the previous day, between when children got up and when they went to bed. For Study 3, one of two dietitians interviewed each child individually either in person or by telephone in the evening of the day school meals were observed; the target period for these interviews was that day, between when children got up and the time of the interview.⁵ In no study did a dietitian interview a child about meals for which that dietitian was the observer. Interviewers followed a multiple-pass protocol modeled after the Nutrition Data System for Research (NDSR; version 4.03, Nutrition Coordinating Center, University of Minnesota, Minneapolis, 2000).⁶ Instead of using the NDSR

⁴For school lunch, *offer-versus-serve* is required in high school but optional in junior high and elementary school; for school breakfast, *offer-versus-serve* is optional in all grades (U.S. Department of Agriculture, Food and Nutrition Service, n.d.).

⁵In no study was the interview strictly a 24hDR because children were not asked if they ate during the night. Also, the target period was slightly shorter for Study 3 than for Studies 1 and 2 because it ended at the time of the evening interview (between 6:30 p.m. and 9:00 p.m.) instead of when the interviewed child went to bed.

⁶The NDSR is a computerized interview protocol and nutrient data base that was developed to obtain 24hDRs in research studies (Dennis, Ernst, Hjortland, Tillotson, & Grambsch, 1980; NDSR Nutrition Data System for Research, n.d.), including studies with children (e.g., Lytle et al., 1998; Weber et al., 2004).

computerized version, interviewers wrote information reported by children on paper forms. In each study, the multiple-pass protocol included four phases: obtain quick list, review quick list, obtain detailed information about quick list, and review all information. After the 24hDR was completed, children were asked to rate their liking of each item reported and/or observed eaten as *not at all*, *a little*, or *a lot* (see, e.g., Baxter, Thompson, Davis, & Litaker, 1999). The Appendix exemplifies the prompts and response options used. Children reported amounts eaten in terms of servings. For each study, each interview was audio recorded and transcribed. Training of interviewers was similar for each study and included modeling, practice, and assessment of quality control for interviews (Shaffer et al., 2004). Throughout data collection for each study, quality control for interviews was assessed regularly and indicated that established criteria was met for conducting interviews properly (see the Appendix).

For Study 1, children were prompted to report meals and snacks eaten in forward order. For Study 2, half of the children in each sex/race group were randomly assigned to forward-order prompts during the first interview and reverse-order prompts during the second interview; the other half received the complementary assignment. For Study 3, all interviews used forward-order prompts; half of the children in each sex/race group were randomly assigned to be interviewed in person, and the other half by telephone.

Coding of variables

For this article, we analyzed only each child's first interview in Studies 1 and 2. Because accuracy in Study 3 did not depend on interview modality (Baxter, Thompson, Litaker, et al., 2003), we collapsed the telephone and in-person groups for the analyses in this article. We defined four groups according to whether the interview used forward-order prompts and was conducted in the morning about the previous day's intake in Study 1 (F-MPD₁) or Study 2 (FMPD₂), used reverse-order prompts and was conducted in the morning about the previous day's intake (R-MPD), or used forward-order prompts and was conducted in the evening about that same day's intake (F-ESD).

Children were to report all eating occasions for the specified target period, but assessment of reporting accuracy in each study was restricted to information about school meals because only these meals were observed. For meals reported to be considered school meals, children had to identify *school* as the location where meals were eaten, refer to breakfast as *breakfast* or *school breakfast* and to lunch as *lunch* or *school lunch*, and report mealtimes to within an hour of observed mealtimes. According to these criteria, eight of 294 children from the three studies (1 F-MPD₂, 6 R-MPD, 1 in-person F-ESD) failed to report both school meals.⁷ Data from these eight children were not included in the analyses for this article.

For each school meal, each item *reported* eaten in any amount more than none was classified as a *match* if it was observed eaten in any amount more than none at the meal, or an *intrusion* if it was not observed eaten at the meal. Each intrusion was further classified as a *stretch* if it was observed on the child's tray for that meal but the amount observed eaten was none, or as a *confabulation* if it was not observed on the child's tray for that meal. Items mentioned but for which amounts reported eaten were *none* were not counted as reported items. Each item *observed* eaten in any amount more than none at the meal was classified as an *omission* if it was not reported eaten at the meal. Because children can report foods many ways, items reported eaten were classified as matches unless it was clear that reports did not describe items observed eaten; this may have overclassified reported items as matches and

⁷Of the six criterial pieces of information (meal name, time, and location for breakfast, and for lunch), these eight children correctly reported an average of 2.25 pieces (range = 1 to 4)—an average of one piece (range = 0 to 2) for breakfast, and 1.25 pieces (range = 0 to 2) for lunch. For breakfast, reporting deficiencies included meal name for six children, location for six children, and time for four children. For lunch, deficiencies included time for seven children, location for five children, and meal name for two children.

underclassified reported items as intrusions. The original coding was conducted by the first author and several dietitians and then checked by computer programming. For this article, another person not involved in any of the three studies coded a random 20% of interviews from each study. Kappa coefficients between the original coding team and the new coder exceeded 0.97 for each of matches, intrusions, stretches, confabulations, and omissions.

Additional details about the samples, observations, interviews, and coding of reported and/or observed information is found in the primary articles about these studies (Baxter et al., 2002; Baxter, Thompson, Litaker, et al., 2003; Baxter, Thompson, Smith, et al., 2003).

Analyses

A logistic regression model was fit to investigate the significance of interview time/target period and reporting-order prompts on the likelihood of each of four outcome variables. Each of three models (concerning reported items that were intrusions, confabulations, and stretches) included one statistical observation (i.e., data row or contribution to the likelihood) for each reported food item. Across the 286 children who reported at least one school meal, there were 1,782 reported items with 1,190 matches and 592 intrusions. Of the 592 intrusions, 119 were stretches and 473 were confabulations. One model investigated the likelihood that a reported item was an intrusion (592/1,782). A second model investigated the likelihood that a reported item was a confabulation (473/1,782). A third model investigated the likelihood that a reported item was a stretch (119/1,782). One additional model investigated the likelihood that an intrusion was a confabulation (473/592).⁸ For each model, race and sex were included, and the standard errors were adjusted to address the fact that statistical observations were clustered within interview (child) (see Williams, 2000).

For the subset of 264 children who had at least one uneaten item that had the potential to be reported, and therefore to be classified as a stretch, we fit a binomial model to investigate the number of stretches out of the number of uneaten items on each child's tray for the two meals combined. This model included the four groups, race, and sex. Standard errors were adjusted to protect for within-child correlation using the sandwich variance estimator.

In addition, for each of these 264 children, we calculated a discriminability measure over the two meals as $[(\text{number of matches}) \div (\text{number of observed items})] - [(\text{number of stretches}) \div (\text{number of uneaten items on the tray})]$. We analyzed this measure using analysis of variance with the four groups as well as race and sex in the model.

To investigate liking ratings, we estimated an ordered logistic regression model for the liking ratings on indicator variables for the four groups and item classifications (match, omission, stretch, confabulation); race and sex were included in the model. In this and subsequent models, standard errors were calculated from a modified sandwich variance estimator accounting for multiple statistical observations per child. The proportional odds assumption for the ordered logistic regression model was rejected [$\chi^2(8) = 81.95, p < 0.001$], indicating a need to investigate a partial proportional odds model. The final model was a generalized ordered logistic regression model for which the coefficient vector was partially restricted to be the same (coefficients on race, sex, F-MPD₁, and R-MPD were restricted) when considering the two comparisons in the outcome variable (*not at all* versus *a little*; *a little* versus *a lot*).

⁸An intrusion could only be classified as a stretch if the item to which it referred was observed on the child's tray for that meal and the amount observed eaten was none. Thus, to determine likelihoods conditionalized on the possibility of reporting a stretch, the logistic regression models were run using data only from the subset of children who could have had a stretch. Across the 264 children in the subset (101 F-MPD₁, 49 F-MPD₂, 54 R-MPD, 60 F-ESD), there were 1,641 reported items with 1,079 matches and 562 intrusions, with 119 stretches and 443 confabulations. Results essentially agreed with those from the full sample of 286 children, so the latter are described in the main text. Exceptions are indicated in footnotes.

We conducted crossmeal examinations to determine whether intrusions that occurred during reports of school breakfast referred to items observed eaten at school lunch, and vice versa. Furthermore, we expanded crossmeal examinations to items observed on the child's tray at the other school meal but for which the amount observed eaten was none.

Analyses were conducted using Stata (9.2, Stata, Inc., College Station, TX, USA) and SAS (8.02, SAS Institute, Inc., Cary, NC, USA).⁹

RESULTS

Table 1 shows, for each group, the numbers of items observed eaten and numbers of items reported eaten for the two school meals. Analysis of variance revealed no significant variation among groups in the numbers of items observed eaten at school meals ($p = 0.5946$), but did detect significant variation among groups in the numbers of items children reported eating for school meals ($p = 0.0089$). Post-hoc two-sample t tests indicated that significantly more items were reported eaten for F-MPD₁ interviews than for R-MPD interviews ($p = 0.0035$) and for F-ESD interviews than for R-MPD interviews ($p = 0.0060$).

Table 2 shows, for each group, the numbers of stretches and confabulations by meal component. For each group, there were more confabulations than stretches for most meal components. Beverages had large percentages of both stretches (37% to 50%) and confabulations (23% to 43%).

Likelihood of intrusions, confabulations, and stretches

Table 3 shows predicted probabilities from logistic regression models of the likelihood that reported items were intrusions, confabulations, and stretches, and the likelihood that intrusions were confabulations, for each of the four groups.

The likelihood that a reported item was an intrusion was significantly higher for F-MPD₁ interviews (odds ratio [OR] = 3.33, $p < 0.001$), F-MPD₂ interviews (OR = 2.68, $p < 0.001$), and R-MPD interviews (OR = 1.92, $p = 0.002$) than for F-ESD interviews. Post-hoc contrasts indicated that the likelihood that a reported item was an intrusion was significantly higher for F-MPD₁ interviews than for R-MPD interviews ($p = 0.0102$).

The likelihood that a reported item was a confabulation was significantly higher for F-MPD₁ interviews (OR = 4.53, $p < 0.001$), F-MPD₂ interviews (OR = 3.72, $p < 0.001$), and R-MPD interviews (OR = 2.55, $p < 0.001$) than for F-ESD interviews. Post-hoc contrasts indicated that the likelihood that a reported item was a confabulation was significantly higher for F-MPD₁ interviews than for R-MPD interviews ($p = 0.0165$).

The likelihood that a reported item was a stretch did not vary significantly among groups.¹⁰

The likelihood that an intrusion was a confabulation was significantly higher for F-MPD₁ interviews (OR = 3.55, $p < 0.001$), F-MPD₂ interviews (OR = 3.42, $p = 0.001$), and R-MPD interviews (OR = 2.58, $p = 0.011$) than for F-ESD interviews. Furthermore, the likelihood that

⁹In the original publications of the three studies (Baxter et al., 2002; Baxter, Thompson, Litaker, et al., 2003; Baxter, Thompson, Smith, et al., 2003), items observed eaten, and reported eaten, were weighted by meal component (combination entrée = 2, condiment = 0.33, remaining meal components = 1) so errors for condiments counted less than errors for most meal components, and errors for combination entrées counted more. The analyses for this article were conducted using unweighted items and meal-component-weighted items; because results essentially agreed, only those for unweighted items are described in the text. Exceptions are indicated in footnotes.

¹⁰Exceptions: For meal-component-weighted items, the likelihood that a reported item was a stretch was significantly lower for White children than for Black children for the full sample of 286 children (OR = 0.60, $p = 0.030$) as well as for the subset of 264 children (OR = 0.63, $p = 0.047$).

an intrusion was a confabulation was significantly higher for White than for Black children (OR = 1.68, $p = 0.032$).¹¹ Post-hoc contrasts failed to indicate any other significant differences.

Stretches and discriminability measure

Of 264 children (132 Black) who could have reported a stretch, 88 children did. Fifty Black children reported stretches; each of 34, 11, 4, and 1 child(ren) reported 1, 2, 3, and 4 stretch(es), respectively. Thirty-eight White children reported stretches; each of 33, 3, 1, and 1 child(ren) reported 1, 2, 3, and 5 stretch(es), respectively.

The proportion of uneaten items on the child's tray reported as stretches for the two meals combined was significantly smaller for White children than for Black children (OR = 0.63, $p = 0.043$) with respective means (and standard deviations [SD]) of 0.11 (0.20) and 0.17 (0.27). No other effects were significant.

For the discriminability measure, means (SD) were 0.33 (0.24) for F-MPD₁ interviews, 0.31 (0.31) for F-MPD₂ interviews, 0.35 (0.29) for R-MPD interviews, and 0.46 (0.42) for F-ESD interviews.¹² A contrast comparing F-ESD to the other three groups showed that the F-ESD mean was significantly larger ($p = 0.0418$).

Liking ratings

Table 4 shows descriptive information concerning liking ratings for each of the four groups by item classification (match, omission, stretch, confabulation). Because the proportional odds assumption did not hold for the analysis of liking ratings, comparison of *not at all* to *a little*, and of *a little* to *a lot*, must be interpreted separately. In comparing *not at all* to *a little*, confabulations (OR = 4.4, $p < 0.001$), stretches (OR = 2.6, $p = 0.012$), and matches (OR = 8.1, $p < 0.001$) were all more likely than omissions to be rated as liked *a little* than *not at all*; furthermore, boys were more likely than girls to rate items as liked *a little* than *not at all* (OR = 1.3, $p = 0.021$). In comparing *a little* to *a lot*, confabulations (OR = 1.5, $p = 0.001$) and matches (OR = 1.9, $p < 0.001$) were more likely than omissions to be rated as liked *a lot* than *a little*, while stretches (OR = 0.6, $p = 0.017$) were less likely; again, boys were more likely than girls to rate items as liked *a lot* than *a little* (OR = 1.3, $p = 0.021$). None of race, group indicators, or any two-way interactions between item classification and group was significant.

Crossmeal examinations

In reports of breakfast, there were 31 intrusions that referred to items observed eaten at lunch by 14 girls and 17 boys, with 12, 6, 5, and 8 from F-MPD₁, F-MPD₂, R-MPD, and F-ESD, respectively; when we expanded this to intrusions that referred to items on the child's tray at lunch but uneaten, there was 1 additional item by a boy for F-MPD₁. In reports of lunch, there were 25 intrusions that referred to items observed eaten at breakfast by 12 girls and 10 boys, with 7, 9, 4, and 5 from F-MPD₁, F-MPD₂, R-MPD, and F-ESD, respectively; one girl [F-MPD₁] had 2 intrusions, and two boys [R-MPD; F-ESD] had 2 intrusions each. When we expanded this to intrusions that referred to items on the child's tray at breakfast but uneaten, there were 6 additional items by 1 girl and 5 boys with 4 from F-MPD₁, 1 from F-MPD₂, and 1 from F-ESD.¹³

¹¹Exception: For unweighted items for the subset of 264 children, the effect of race was not significant.

¹²Values may range from -1 to 1; values closer to 1 indicate that for the items reported eaten, there was better ability to discriminate between what was eaten and what could have been eaten but was not.

DISCUSSION

Our goal was to enhance understanding of intrusions in children's dietary recalls by examining the occurrence of intrusions and two types of intrusion, as well as the role of liking, at different retention intervals and under different reporting-order prompts.

Both the occurrence of intrusions and types of intrusion were related to target period and interview time, the variables that, together, were a proxy for retention interval. As the retention interval increased (from evening interviews about the same day's intake to morning interviews about the previous day's intake), the likelihood that reported items were intrusions increased 1.92 to 3.33 times. This finding, which is consistent with other results that show that reporting accuracy decreases as retention interval increases, is compatible with both the model of dietary reporting proposed by Smith and Jobe (1994) and the fuzzy trace theory of Reyna and Brainerd (1995). Moreover, the types of intrusion changed with retention interval: As the retention interval increased, the likelihood that reported items were confabulations increased 2.55 to 4.53 times, and the likelihood that intrusions were confabulations increased 2.58 to 3.55 times. These results, too, are consistent with both the Smith and Jobe model and the fuzzy trace theory. However, the likelihood that reported items were stretches did not vary over retention intervals.

Concerning reporting-order prompts, although the likelihoods that reported items were intrusions and that reported items were confabulations were both greater with forward-order prompts in Study 1 than with reverse-order prompts in Study 2, significant differences were not found between the forward- and reverse-order prompts groups of Study 2. In other words, the effect of reporting-order prompts was significant across studies but not within study. Whether and how reporting-order prompts affect the likelihood and types of children's dietary intrusion is, at best, unresolved. Note, first, that no significant difference was found in the likelihood that a reported item was an intrusion with forward-order prompts between Studies 1 and 2. Second, although the number of items observed eaten at school meals did not differ significantly among the groups, fewer items were reported eaten in R-MPD interviews than in F-MPD₁ and F-ESD interviews. Finally, six of the eight children who were classified as reporting no school meals (and not included in these analyses because they contributed no reported items and thus no intrusions) were in the reverse-order condition of Study 2. Thus, compared to forward-order prompts, reverse-order prompts may have had some effect on reporting performance in general, and on the likelihood of intrusions in particular, that was not detected by these analyses. No predictions concerning reporting-order prompts were based on the model of Smith and Jobe (1994); a weak case was made that fuzzy trace theory would predict fewer intrusions with reverse-order prompts than with forward-order prompts.

Items on a child's tray that were eaten, items on a child's tray that were uneaten, and the universe of other items may be conceptualized as three sources of reportable items; only items from the first of these sources should be reported. The first two sets are finite; each item in each of these two sets was either reported by the child or not, so the difference of probabilities of reporting items from these sets quantifies the discriminability of these sets. This measure was significantly larger for F-ESD interviews than for the average of the previous day interviews, indicating that the discrimination between to-be-reported eaten items and not-to-be-reported uneaten items was more effective by children at the shorter than at the longer retention interval. Not surprisingly, confabulations outnumbered stretches. Vastly more items had the potential

¹³For reports of breakfast, the 31 items consisted of 25 milks (8 white, 10 chocolate, 7 strawberry), 1 mixed fruit, 1 orange juice, 1 ketchup, 2 pizzas, and 1 biscuit; the additional item was white milk. For reports of lunch, the 25 items consisted of 15 milks (9 white, 5 chocolate, 1 strawberry), 4 juices (3 apple, 1 fruit blend), 1 mixed fruit, 1 peaches, 1 graham crackers, 1 ketchup, and 2 pizzas; the 6 additional items were 4 milks (2 white, 2 chocolate), 1 graham crackers, and 1 pizza. The results for crossmeal examinations include 2 girls (F-MPD₂ and R-MPD) who each had a breakfast intrusion that referred to an item observed eaten at lunch as well as a lunch intrusion that referred to an item observed eaten at breakfast. For each girl, the 2 items were white milk at 1 meal and chocolate milk at the other.

to be confabulations than stretches: Only the limited number of uneaten items on children's trays could be stretches, whereas the set of potential confabulations was almost unlimited.

Baxter et al. (1999), in a study with fourth-grade children, found that for items liked *a lot*, omission rates were lower for next morning lunch recalls, and intrusion rates were lower for both 1.5 hour and next morning lunch recalls, than for items *not liked a lot*. (Types of intrusion were not distinguished in that study.) This finding suggested that children report items that they like, and that liked items, although uneaten, might be reported. In the studies described in this article, most items were served to children; thus, liking did not influence whether most items were *on* children's school meal trays, but could have influenced *which* items children ate. Liking ratings were higher for matches than stretches, for confabulations than stretches, and for matches than omissions, as predicted in the Introduction. For the current analyses of school meals reported during 24hDRs, none of the group indicators for retention interval or reporting-order prompts were significant in the analyses of liking ratings.

Crossmeal examinations indicated that only 9.5% of the 592 intrusions referred to items eaten at the other school meal; furthermore, expanding the crossmeal examination to uneaten items on the child's tray at the other school meal accounted for only an additional 1.2% of intrusions. These results suggest that confusing items eaten across school meals, and confusing items available but uneaten across school meals, are not major sources of fourth-grade children's intrusions in reports of school meals.

In a few of the analyses, race or sex was significant. These findings were unexpected, and we have no explanations for them.

This investigation has several limitations. The data analyzed were from three studies, each of which had its own purpose and hypothesis; none was designed to investigate the issues addressed in this article. It would have been desirable to address these issues in a single study designed for this purpose. Because conducting dietary validation studies is expensive and time consuming, we utilized data previously collected to address these issues, and stimulate future research on this topic. As indicated in the Appendix, the prompts and response options to obtain information about meal names, locations, and liking ratings varied slightly across the three studies. Because subjects for all three studies were fourth-grade children, developmental trends concerning stretches and confabulations in children's 24hDRs could not be examined. Future validation studies across an age range in childhood would be powerful in both an empirical and theoretical sense, and would permit comparison of accurate and inaccurate memory for school meals with studies concerning false memory for other stimulus materials. None of the three studies provides baseline information about children's ability to report, during 24hDRs, items eaten at two school meals; baseline information is available only from ten children in the study by Baxter et al. (2004).¹⁴

Several strengths help offset these limitations. School meal observations were used to validate these parts of children's 24hDRs. For each study, the sex/race profile of children who provided child assent and parental consent to participate was similar to that of the total population of fourth-grade children at the schools from which they were recruited, which suggests that the sample for each study was representative of the population. For each study, considerably more children agreed to participate (with their parent's consent) than were needed, which made it difficult for individual children to know who was being observed and would be interviewed.

¹⁴Results from a retrospective analysis using data from two validation studies indicated that fourth-grade children's omission rates and intrusion rates for every meal component were greater (worse) for the school lunch part of 24hDRs than for recalls of only school lunch (Baxter & Thompson, 2002). Thus, the cognitive burden of recalling items eaten at school lunch as part of a 24hDR appears to be greater than that of recalling school lunch items as a single meal, and negatively impacts fourth-grade children's ability to report their dietary intake.

Quality control for observations and interviews, assessed regularly throughout data collection for each study, was acceptable.

It has long been known that as the retention interval increases between relatively mundane behaviors and reports about those behaviors, reporting accuracy deteriorates. The current article provides some details about the nature of that deterioration in fourth-grade children's reports of school meals obtained during 24hDRs. The practical significance of this line of research is twofold. First, because most studies that rely on 24hDRs lack a validation method, most studies analyze intrusions along with matches when assessing dietary intake. Second, the prevalence of overweight children in the United States has increased (Centers for Disease Control and Prevention, n.d.), which makes accurate assessment of children's dietary intake a national health priority. Our results suggest the need for increased attention to retention interval and reporting-order prompts in studies that rely on children's 24hDRs. Dietary validation studies designed to investigate types of intrusion may help guide the development of specific recommendations to reduce the occurrence of intrusions, and thus improve children's dietary recall accuracy.

Acknowledgments

Data collection was supported by grant R01HL63189 from the National Heart, Lung, and Blood Institute (NHLBI) of the National Institutes of Health (NIH). The additional analyses and manuscript preparation were supported by grant R01HL73081 from NHLBI of NIH. Suzanne D. Baxter was Principal Investigator (PI) for both grants. Albert F. Smith was a Consultant on grant R01HL63189 and PI on a sub-contract to grant R01HL73081.

We thank the children and staff of Blythe, Goshen, Gracewood, Hephzibah, Lake Forest Hills, McBean, Monte Sano, National Hills, Rollins, Southside, Willis Foreman, and Windsor Spring Elementary Schools, the School Nutrition Program, and the Richmond County (Georgia) Board of Education for allowing data collection. We thank Alyssa Mackelprang for serving as the new coder for the current article.

Contract/grant sponsor: National Heart, Lung, and Blood Institute of the National Institutes of Health; contract/grant numbers: R01HL63189 and R01HL73081.

Appendix

Details by study relevant to the current analyses

	Study 1 (Baxter et al., 2002)	Study 2 (Baxter, Thompson, Smith, et al., 2003)	Study 3 (Baxter, Thompson, Litaker, et al., 2003)
School year of data collection	1999–2000	2000–2001	2001–2002
Number of schools	6	11	10
Percent of children by school (most to least) eligible for free or reduced-price school meals during the school year of data collection	91%, 83%, 71%, 68%, 66%, 36%	74%, 73%, 72%, 69%, 66%, 64%, 59%, 57%, 50%, 40%, 33%	79%, 72%, 71%, 67%, 61%, 60%, 59%, 46%, 45%, 27%
Number (%) of children who agreed to participate	382 (73%)	669 (73%)	451 (56%)
Inter-observer reliability – frequency throughout data collection; total number of children; mean agreement across observers for food items in which amounts observed eaten were within ¼ serving	Twice monthly; 18 children; 89%	Weekly; 22 children; 90%	Weekly; 10 children; 93%
Number of children interviewed	104	121	69
Interviews per child	1 (n = 12), 2 (n = 13), 3 (n = 79) with > 28 days between any 2 interviews	2 separated by ≥ 29 days	1
Quality control for interviews (QCI) – frequency throughout data collection	Monthly ¹	Weekly ²	Daily ³

	Study 1 (Baxter et al., 2002)	Study 2 (Baxter, Thompson, Smith, et al., 2003)	Study 3 (Baxter, Thompson, Litaker, et al., 2003)
Interview time	Morning (after breakfast but before lunch)	Morning (after breakfast but before lunch)	Evening (between 6:30 p.m. and 9:00 p.m.)
Target period	Previous day's intake	Previous day's intake	Same day's intake
Reporting-order prompts for meals and snacks eaten	Forward	Forward for one interview and reverse for the other	Forward
Interview modality	In person	In person	Telephone (n = 33) or in person (n = 36)
The name of each meal reported was obtained by asking the child:	"What was the name of the meal you ate at [time]?" (Response options shown on a cue card were school breakfast, breakfast, school lunch, lunch, dinner/supper, snack, other.)	"What was the name of the meal you ate at [time] yesterday [morning, afternoon, evening] when you had [list food/drinks]? Please choose from these: school breakfast, breakfast, school lunch, lunch, dinner/supper, snack, other." (Response options were shown on a cue card and read aloud.)	"What was the name of the meal you ate at [time this morning, today, this afternoon, tonight] when you had [list food/drinks]? Please listen to all of these choices and choose one: school breakfast, breakfast, school lunch, lunch, dinner/supper, or snack." (No cue card was used.)
The location where each meal was eaten was obtained by asking the child:	"Where did you eat this?" (Response options shown on a cue card were home, friend's home, school, day care, restaurant/fast food, deli/take-out store, community meal program, party/reception/sporting event, other.)	"Where did you eat this? Please choose from these: home, school, other." (Response options were shown on a cue card and read aloud.)	"Where did you eat this? Please choose from these: home, school, or somewhere else." (No cue card was used.)
Towards the end of the interview after the dietary recall was completed, a liking rating for each item observed on the child's tray, and/or reported eaten by the child, for each school meal was obtained by asking the child:	I'd like to know how much you liked the following foods from yesterday. When I read each food to you, please tell me if you liked it a lot, a little, or not at all... [food] at [meal] yesterday?" (One meal was completed before moving to the next. Response options were shown on a cue card.)	I'd like to know how much you liked the following foods from yesterday. When I read each food to you, please tell me if you liked it a lot, a little, or not at all. Starting with [meal] at [time] yesterday [morning, afternoon, evening], how much did you like the ____?" For other items at that meal, only the name of each item was given. One meal was completed before moving to the next. Response options were shown on a cue card and read aloud.)	"I'd like to know how much you liked the following foods and drinks from today. When I read each food to you, please tell me if you liked it a lot, a little, or not at all. Starting with [meal] at [time-this morning, today, this afternoon, tonight] at [location], how much did you like the ____?" (For other items at that meal, only the name of each item was given. One meal was completed before moving to the next. (No cue card was used.)
Omission rates ⁴ (means weighted by meal component) ⁵	51% across 275 recalls; considerable variability from one recall to next recall for same child	57% and 56% with reverse and forward-order prompts; order by sex interaction ($p < 0.008$) – lower for boys with reverse- (53%) versus forward-order (62%) prompts but not for girls (61% versus 53%)	34% for in person recalls; 32% for telephone recalls
Intrusion rates ⁶ (means weighted by meal component) ⁵	39% across 275 recalls; considerable variability from one recall to next recall for same child	32% and 39% with reverse- and forward-order prompts; $\leq 30\%$ for boys for 54% and 40% of recalls with reverse and forward-order prompts ($p = 0.095$)	19% for in person recalls; 16% for telephone recalls
Specific interview per child included in the current analyses; number of children (Black girls, Black boys, White girls, White boys)	First interview; 104 children (27, 24, 28, 25)	First interview; forward-order prompts 55 children (13, 14, 13, 15); reverse-order prompts 59 children (14, 13, 16, 16)	Only interview; 68 children (19, 17, 16, 16)

¹ For inter-reviewer reliability (IRR), one dietitian interviewed a child while the other dietitian(s) sat behind the child and completed an interview form according to what was heard. At the end of the interview, the listening dietitian(s) could ask the child questions. Results on 8 children indicated that only one additional question would have been asked if a different dietitian had conducted the interview. (These 8 interviews were used for IRR only.) In addition, the Principal Investigator (PI) assessed QCI by reviewing the interview form, audio recording, and typed transcript for a random 5% of each dietitian's interviews and verified adherence to the interview protocol across dietitians.

² The interview form, audio recording, and typed transcript from one interview conducted each week by each interviewer was randomly selected and reviewed by another interviewer who completed a QCI checklist by marking *adequate*, *needs improvement*, or *not applicable* for each of 53 items. Each QCI checklist was reviewed by the PI and interviewer. Across 51 QCI checklists (21% of the 242 interviews), 95% of items (2,133 of 2,249) that were not rated *not applicable* were rated *adequate*; this indicates that interviewers adequately followed the interview protocols.

³ The interview form, audio recording, and typed transcript from one interview conducted each evening was randomly selected and reviewed by another dietitian, who completed a QCI checklist by marking *adequate*, *needs improvement*, or *not applicable* for each of 56 items. Each QCI checklist was reviewed by the PI and interviewer. Across 22 QCI checklists (32% of the 69 interviews), 97% of items (937 of 970) that were not rated *not applicable* were rated *adequate*; this indicates that interviewers adequately followed the interview protocols.

⁴ Omission rate = (omissions / omissions + matches) $\times$ 100; it is the proportion of items observed eaten but not reported eaten. Values range from 0% to 100%, with higher rates indicating more omissions. 100% – omission rate = Match rate.

⁵ In the original publications of the three validation studies, items were given meal-component weights (combination entrée = 2, condiment = 0.33, remaining meal components = 1) so that errors for condiments (e.g., ketchup) counted less than errors for most meal components, and errors for combination entrées (e.g., cheeseburger) counted more.

⁶Intrusion rate = (intrusions / intrusions + matches) × 100; it is the proportion of items reported eaten but not observed eaten. Values range from 0% to 100%, with higher rates indicating more intrusions.

REFERENCES

- Armstrong AM, MacDonald A, Booth IW, Platts RG, Knibb RC, Booth DA. Errors in memory for dietary intake and their reduction. *Applied Cognitive Psychology* 2000;14:183–191. <183:AID-ACP645>3.0.CO;2-#
- Baglio ML, Baxter SD, Guinn CH, Thompson WO, Shaffer NM, Frye FHA. Assessment of interobserver reliability in nutrition studies that use direct observation of school meals. *Journal of the American Dietetic Association* 2004;104:1385–1392. [PubMed: 15354155]
- Baranowski T, Dworkin R, Henske JC, Clearman DR, Dunn JK, Nader PR, et al. The accuracy of children's self-reports of diet: Family Health Project. *Journal of the American Dietetic Association* 1986;86:1381–1385. [PubMed: 3760429]
- Baranowski T, Islam N, Baranowski J, Cullen KW, Myres D, Marsh T, et al. The Food Intake Recording Software System is valid among fourth-grade children. *Journal of the American Dietetic Association* 2002;102:380–385. [PubMed: 11902371]
- Baranowski T, Simons-Morton BG. Dietary and physical activity assessment in school-aged children: Measurement issues. *Journal of School Health* 1991;61:195–197. [PubMed: 1943040]
- Baxter SD, Smith AF, Guinn CH, Thompson WO, Litaker MS, Baglio ML, et al. Interview format influences the accuracy of children's dietary recalls validated with observations. *Nutrition Research* 2003;23:1537–1546. [PubMed: 16724161]
- Baxter SD, Smith AF, Hardin JW, Nichols MN. Conventional energy and macronutrient variables distort the accuracy of children's dietary reports: Illustrative data from a validation study of effect of order prompts. *Preventive Medicine* 2007;44:34–41. [PubMed: 16959308]
- Baxter SD, Smith AF, Litaker MS, Guinn CH, Shaffer NM, Baglio ML, et al. Recency affects reporting accuracy of children's dietary recalls. *Annals of Epidemiology* 2004;14:385–390. [PubMed: 15246326]
- Baxter SD, Thompson WO. Accuracy by meal component of fourth-graders' school lunch recalls is less when obtained during a 24-hour recall than as a single meal. *Nutrition Research* 2002;22:679–684.
- Baxter SD, Thompson WO, Davis HC. Trading of food during school lunch by first- and fourth-grade children. *Nutrition Research* 2001;21:499–503.
- Baxter SD, Thompson WO, Davis HC, Johnson MH. Impact of gender, ethnicity, meal component, and time interval between eating and reporting on accuracy of fourth-graders' self-reports of school lunch. *Journal of the American Dietetic Association* 1997;97:1293–1298. [PubMed: 9366868]
- Baxter SD, Thompson WO, Davis HC, Litaker MS. Children's dietary recalls: The salience of entrée and liking for foods on accuracy and order of reporting. *Nutrition* 1999;15:848–853. [PubMed: 10575659]
- Baxter SD, Thompson WO, Litaker MS, Frye FHA, Guinn CH. Low accuracy and low consistency of fourth-graders' school breakfast and school lunch recalls. *Journal of the American Dietetic Association* 2002;102:386–395. [PubMed: 11905461]
- Baxter SD, Thompson WO, Litaker MS, Guinn CH, Frye FHA, Baglio ML, et al. Accuracy of fourth-graders' dietary recalls of school breakfast and school lunch validated with observations: In-person versus telephone interviews. *Journal of Nutrition Education and Behavior* 2003;35:124–134. [PubMed: 12773283]
- Baxter SD, Thompson WO, Smith AF, Litaker MS, Yin Z, Frye FHA, et al. Reverse versus forward order reporting and the accuracy of fourth-graders' recalls of school breakfast and school lunch. *Preventive Medicine* 2003;36:601–614. [PubMed: 12689806]
- Brainerd CJ, Reyna VF. Fuzzy-trace theory and memory development. *Developmental Review* 2004;24:396–439.
- Brainerd, CJ.; Reyna, VF. *The Science of False Memory*. New York: Oxford University Press; 2005.
- Buzzard, M. 24-hour dietary recall and food record methods. In: Willett, W., editor. *Nutritional Epidemiology*. Vol. 2nd ed.. New York, NY: Oxford University Press; 1998. p. 50-73.

- Caballero B, Davis S, Davis CE, Ethelbah B, Evans M, Lohman T, et al. A school-based program for the primary prevention of obesity in American Indian children. *Journal of Nutritional Biochemistry* 1998;9:535–543.
- Centers for Disease Control and Prevention (n.d.). Prevalence of overweight among children and adolescents United States 1999–2002. Retrieved August 8, 2007 from <http://www.cdc.gov/nchs/products/pubs/pubd/hestats/overwght99.htm>
- Contento IR, Randell JS, Basch CE. Review and analysis of evaluation measures used in nutrition intervention research. *Journal of Nutrition Education and Behavior* 2002;34:2–25. [PubMed: 11917668]
- Crawford PB, Obarzanek E, Morrison J, Sabry ZI. Comparative advantage of 3-day food records over 24-hour recall and 5-day food frequency validated by observation of 9- and 10-year-old girls. *Journal of the American Dietetic Association* 1994;94:626–630. [PubMed: 8195550]
- Davis MR, McMahon M, Greenwood KM. The efficacy of mnemonic components of the cognitive interview: Towards a shortened variant for time-critical investigations. *Applied Cognitive Psychology* 2005;19:75–93.
- Dennis B, Ernst N, Hjortland M, Tillotson J, Grambsch V. The NHLBI Nutrition Data System. *Journal of the American Dietetic Association* 1980;77:641–647. [PubMed: 6893713]
- Eck LH, Klesges RC, Hanson CL. Recall of a child's intake from one meal: Are parents accurate? *Journal of the American Dietetic Association* 1989;89:784–789. [PubMed: 2723300]
- Emmons L, Hayes M. Accuracy of 24-hr. recalls of young children. *Journal of the American Dietetic Association* 1973;62:409–415. [PubMed: 4691953]
- Field AE, Peterson KE, Gortmaker SL, Cheung L, Rockett H, Fox MK, et al. Reproducibility and validity of a food frequency questionnaire among fourth to seventh grade inner-city school children: Implications of age and day-to-day variation in dietary intake. *Public Health Nutrition* 1999;2:293–300. [PubMed: 10512564]
- Frank GC. Taking a bite out of eating behavior: Food records and food recalls of children. *Journal of School Health* 1991;61:198–200. [PubMed: 1943041]
- Fries E, Green P, Bowen DJ. What did I eat yesterday? Determinants of accuracy in 24-hour food memories. *Applied Cognitive Psychology* 1995;9:143–155.
- Geiselman RE, Callot R. Reverse versus forward recall of script-based texts. *Applied Cognitive Psychology* 1990;4:141–144.
- Gleason, P.; Suitor, C. *Children's Diets in the Mid-1900s: Dietary Intake and Its Relationship with School Meal Participation. CN-01-CD1* (No. CN-01-CD1). Alexandria, VA: US Department of Agriculture, Food and Nutrition Service, Office of Analysis, Nutrition and Evaluation, Project Officer, Ed Herzog; 2001. Retrieved August 8, 2007 from <http://www.fns.usda.gov/oane/menu/published/CNP/FILES/ChilDiet.pdf>
- Jensen JK, Gustafson D, Boushey CJ, Auld G, Bock MA, Bruhn CM, et al. Development of a food frequency questionnaire to estimate calcium intake of Asian, Hispanic, and White youth. *Journal of the American Dietetic Association* 2004;104:762–769. [PubMed: 15127061]
- Jobe JB, White AA, Kelley CL, Mingay DJ, Sanchez MJ, Loftus EF. Recall strategies and memory for health care visits. *Milbank Memorial Fund Quarterly* 1990;68:171–189.
- Jones CH, Pipe M-E. How quickly do children forget events? A systematic study of children's event reports as a function of delay. *Applied Cognitive Psychology* 2002;16:755–768.
- Koriat A, Goldsmith M. Memory in naturalistic and laboratory contexts: Distinguishing the accuracy-oriented and quantity-oriented approaches to memory assessment. *Journal of Experimental Psychology: General* 1994;123:297–315. [PubMed: 7931094]
- Koriat A, Goldsmith M. Memory metaphors and the real-life/laboratory controversy: Correspondence versus storehouse conceptions of memory. *Behavioral and Brain Sciences* 1996;19:167–228.
- Koriat A, Goldsmith M, Pansky A. Toward a psychology of memory accuracy. *Annual Review of Psychology* 2000;51:481–537.
- Loftus, EF.; Smith, KD.; Klinger, MR.; Fiedler, J. Memory and mismemory for health events. In: Tanur, JM., editor. *Questions about Questions: Inquiries into the Cognitive Bases of Surveys*. New York: Sage; 1992. p. 102-137.

- Luepker RV, Perry CL, McKinlay SM, Nader PR, Parcel GS, Stone EJ, et al. Outcomes of a field trial to improve children's dietary patterns and physical activity: The Child and Adolescent Trial for Cardiovascular Health (CATCH). *Journal of the American Medical Association* 1996;275:768–776. [PubMed: 8598593]
- Lytle LA, Murray DM, Perry CL, Eldridge AL. Validating fourth-grade students' self-report of dietary intake: Results from the 5-A-Day Power Plus program. *Journal of the American Dietetic Association* 1998;98:570–572. [PubMed: 9597031]
- Lytle LA, Nichaman MZ, Obarzanek E, Glovsky E, Montgomery D, Nicklas T, et al. Validation of 24-hour recalls assisted by food records in third-grade children. *Journal of the American Dietetic Association* 1993;93:1431–1436. [PubMed: 8245378]
- Memon A, Wark L, Bull R, Koehnken G. Isolating the effects of the cognitive interview techniques. *British Journal of Psychology* 1997;88:179–197.
- Mertz W. Food intake measurements: Is there a "gold standard"? *Journal of the American Dietetic Association* 1992;92:1463–1465. [PubMed: 1452958]
- Mitchell, KJ.; Johnson, MK. Source monitoring: Attributing mental experience. In: Tulving, E.; Craik, FIM., editors. *The Oxford Handbook of Memory*. New York: Oxford University Press; 2000. p. 179–195.
- Moshfegh, A.; Goldman, J.; Cleveland, L. *What We Eat in America*, NHANES 2002–2002: Usual nutrient intakes from food compared to Dietary Reference Intakes. U.S. Department of Agriculture, Agricultural Research Service; 2005. Retrieved August 8, 2007 from <http://www.ars.usda.gov/SP2UserFiles/Place/12355000/pdf/usualintaketables2001-02.pdf>
- NDSR Nutrition Data System for Research (n.d.). Retrieved August 8, 2007 from <http://www.ncc.umn.edu/products/ndsr.html>
- Nicklas TA, Webber LS, Srinivasan SR, Berenson GS. Secular trends in dietary intakes and cardiovascular risk factors of 10-y-old children: The Bogalusa Heart Study (1973–1988). *American Journal of Clinical Nutrition* 1993;57:930–937. [PubMed: 8503364]
- Raaijmakers JGW, Shiffrin RM. SAM: A theory of probabilistic search of associative memory. *Psychology of Learning and Motivation* 1980;14:207–262.
- Raaijmakers JGW, Shiffrin RM. Search of associative memory. *Psychological Review* 1981;88:93–134.
- Reyna VF, Brainerd CJ. Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences* 1995;7:1–75.
- Roediger, HL.; Balota, DL.; Watson, JM. Spreading activation and arousal of false memories. In: Roediger, HL.; Nairne, J.; Neath, I.; Surprenant, A., editors. *The nature of remembering: Essays in honor of Robert G. Crowder*. Washington, DC: American Psychological Association; 2001. p. 95–115.
- Samuelson G. An epidemiological study of child health and nutrition in a northern Swedish county. II. Methodological study of the recall technique. *Nutrition & Metabolism* 1970;12:321–340. [PubMed: 5519001]
- Shaffer NM, Baxter SD, Thompson WO, Baglio ML, Guinn CH, Frye FHA. Quality control for interviews to obtain dietary recalls from children for research studies. *Journal of the American Dietetic Association* 2004;104:1577–1585. [PubMed: 15389417]
- Simons-Morton BG, Baranowski T. Observation in assessment of children's dietary practices. *Journal of School Health* 1991;61:204–207. [PubMed: 1943043]
- Simons-Morton BG, Forthofer R, Huang IW, Baranowski T, Reed DB, Fleishman R. Reliability of direct observation of schoolchildren's consumption of bag lunches. *Journal of the American Dietetic Association* 1992;92:219–221. [PubMed: 1737906]
- Smith, AF. *Cognitive Processes in Long-term Dietary Recall*. Hyattsville, MD: National Center for Health Statistics, Vital and Health Statistics; 1991. Series 6, No. 4
- Smith AF. Concerning the suitability of recordkeeping for validating and generalizing about reports of health-related information. *Review of General Psychology* 1999;3:133–150.
- Smith, AF.; Jobe, JB. Validity of reports of long-term dietary memories: Data and a model. In: Schwarz, N.; Sudman, S., editors. *Autobiographical Memory and the Validity of Retrospective Reports*. New York: Springer-Verlag; 1994. p. 121–140.

- Smith AF, Jobe JB, Mingay DJ. Retrieval from memory of dietary information. *Applied Cognitive Psychology* 1991;5:269–296.
- Todd KS, Kretsch MJ. Accuracy of the self-reported dietary recall of new immigrant and refugee children. *Nutrition Research* 1986;6:1031–1043.
- U.S. Department of Agriculture, Food and Nutrition Service (n.d.). Road to SMI Success: A Guide for School Foodservice Directors. [August 8, 2007]. from <http://www.fns.usda.gov/tn/Resources/roadtosuccess.html>
- U.S. Department of Agriculture, Economic Research Service (n.d.). The Food Assistance Landscape, FY 2006 Annual Report. [August 8, 2007]. from <http://www.ers.usda.gov/Publications/EIB6-4>
- Weber JL, Lytle L, Gittelsohn J, Cunningham-Sabo L, Heller K, Anliker JA, et al. Validity of self-reported dietary intake at school meals by American Indian children: The Pathways Study. *Journal of the American Dietetic Association* 2004;104:746–752. [PubMed: 15127059]
- Williams RL. A note on robust variance estimation for cluster-correlated data. *Biometrics* 2000;56:645–646. [PubMed: 10877330]
- Zaromb FM, Howard MW, Dolan ED, Sirotin YB, Tully M, Wingfield A, et al. Temporal associations and prior-list intrusions in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 2006;32:792–804.

Table 1

Means (and standard deviations) of numbers of items observed eaten (in any amount more than none), and numbers of items reported eaten (in any amount more than none), for two school meals (breakfast, lunch), over children in the four groups ^{1,2}

	F-MPD ₁ Study 1 (104 children)	F-MPD ₂ Study 2 (55 children)	R-MPD Study 2 (59 children)	F-ESD Study 3 (68 children)
Number of observed items ³	8.6 (2.1)	8.5 (1.9)	8.1 (2.0)	8.5 (2.5)
Number of reported items ⁴	6.5 (2.1) ^{a,c}	5.9 (2.1) ^{a,b,c}	5.5 (2.1) ^b	6.6 (2.3) ^c

¹ F-MPD interviews used **F**orward-order prompts (morning-to-evening) for meals/snacks eaten and were conducted in the **M**orning about the **P**revious **D**ay's intake; R-MPD interviews used **R**everse-order prompts (evening-to-morning) for meals/snacks eaten and were conducted in the **M**orning about the **P**revious **D**ay's intake; F-ESD interviews used **F**orward-order prompts for meals/snacks eaten and were conducted in the **E**vening about that **S**ame **D**ay's intake.

² Means on the same row with different superscript letters differed significantly.

³ Analysis of variance failed to detect a significant difference in the numbers of items observed eaten between the four groups, $F(3, 282) = 0.63$, $p = 0.5946$.

⁴ Analysis of variance detected a significant difference in the numbers of items reported eaten between the four groups, $F(3, 282) = 3.94$, $p = 0.0089$. Post-hoc two-sample t tests indicated that significantly more items were reported eaten for F-MPD₁ interviews than for R-MPD interviews, $t(161) = 2.9628$, $p = 0.0035$, and for F-ESD interviews than for R-MPD interviews $t(125) = 2.7956$, $p = 0.0060$.

Table 2

Number (and percent) of types of intrusion¹ – stretch and confabulation² – for each of four groups³ by meal component

Meal component	F-MPD ₁ (Study 1) [104 children]				F-MPD ₂ (Study 2) [55 children]				R-MPD (Study 2) [59 children]				F-ESD (Study 3) [68 children]			
	Stretch	Confabulation	Stretch	Confabulation	Stretch	Confabulation	Stretch	Confabulation	Stretch	Confabulation	Stretch	Confabulation	Stretch	Confabulation	Stretch	Confabulation
Beverage	17 (37)	81 (33)	9 (45)	31 (30)	10 (50)	18 (23)	15 (45)	21 (43)								
Bread	9 (20)	58 (24)	4 (20)	21 (21)	2 (10)	23 (29)	4 (12)	9 (18)								
Combination entrée	2 (4)	31 (13)	3 (15)	17 (17)	2 (10)	17 (22)	1 (3)	3 (6)								
Condiment	3 (7)	22 (9)	1 (5)	4 (4)	1 (5)	3 (4)	1 (3)	4 (8)								
Dessert (lunch only)	1 (2)	9 (4)	0 (0)	2 (2)	0 (0)	3 (4)	0 (0)	2 (4)								
Entrée (lunch only)	2 (4)	8 (3)	0 (0)	2 (2)	1 (5)	1 (1)	3 (9)	0 (0)								
Fruit	3 (7)	16 (7)	1 (5)	6 (6)	0 (0)	6 (8)	1 (3)	6 (12)								
Meat (breakfast only)	0 (0)	2 (1)	0 (0)	1 (1)	0 (0)	0 (0)	1 (3)	1 (2)								
Miscellaneous	0 (0)	3 (1)	0 (0)	1 (1)	0 (0)	0 (0)	0 (0)	2 (4)								
Vegetable	9 (20)	14 (6)	2 (10)	17 (17)	4 (20)	7 (9)	7 (21)	1 (2)								
Total n	46	244	20	102	20	78	33	49								

¹ An intrusion is an item that was reported eaten but was not observed eaten at that school meal. Examples of items reported eaten that were classified as intrusions because they differed from items observed eaten included fruit juices (e.g., apple observed, orange reported), milk flavors (e.g., white observed, chocolate reported), dry cereals (e.g., flake-shaped observed, donut-shaped reported), and vegetables (e.g., broccoli observed, green beans reported). Examples of items reported eaten that were classified as matches because they were similar to items observed eaten included all types of pizza (e.g., cheese, pepperoni) and all types of white milk (e.g., whole, 1%, skim).

² An intrusion was further classified either as a *stretch* if it referred to an item that was observed on the child's tray for that meal, but for which the amount observed eaten was none, or as a *confabulation* if it referred to an item that was *not* observed on the child's tray for that meal.

³ F-MPD interviews used **Forward-order prompts** (morning-to-evening) for meals/snacks eaten and were conducted in the **Morning** about the **Previous Day**'s intake; R-MPD interviews used **Reverse-order prompts** (evening-to-morning) for meals/snacks eaten and were conducted in the **Morning** about the **Previous Day**'s intake; F-ESD interviews used **Forward-order prompts** for meals/snacks eaten and were conducted in the **Evening** about that **Same Day**'s intake.

Table 3

Predicted probabilities (and standard errors) from logistic regression models of the likelihood that reported items were intrusions, confabulations, and stretches, and the likelihood that intrusions were confabulations, for each of the four groups ^{1,2,3}

	F-MPD ₁ Study 1 (104 children)	F-MPD ₂ Study 2 (55 children)	R-MPD Study 2 (59 children)	F-ESD Study 3 (68 children)
Intrusions among reported items ⁴	.43 (.03) ^a	.37 (.03) ^{a,b}	.30 (.03) ^b	.18 (.03) ^c
Confabulations among reported items ⁵	.36 (.03) ^d	.31 (.03) ^{d,e}	.24 (.03) ^e	.11 (.02) ^f
Stretches among reported items ⁶	.07 (.02) ^g	.06 (.02) ^g	.06 (.02) ^g	.07 (.02) ^g
Confabulations among intrusions ⁷	.85 (.03) ^h	.84 (.04) ^h	.80 (.05) ^h	.60 (.07) ⁱ

¹ F-MPD interviews used **F**orward-order prompts (morning-to-evening) for meals/snacks eaten and were conducted in the **M**orning about the **P**revious **D**ay's intake; R-MPD interviews used **R**everse-order prompts (evening-to-morning) for meals/snacks eaten and were conducted in the **M**orning about the **P**revious **D**ay's intake; F-ESD interviews used **F**orward-order prompts for meals/snacks eaten and were conducted in the **E**vening about that **S**ame **D**ay's intake. Intrusions were items reported as eaten that had not been observed eaten at that meal. Confabulations were intrusions that were not observed on the child's meal tray. Stretches were intrusions that were observed on the child's meal tray

² For each school meal, observed amounts and reported amounts for each item were recorded in terms of servings of standardized school portions. A reported item was an intrusion if the amount reported eaten was any amount more than none when the observation indicated that none of the item was eaten.

³ Predicted probabilities on the same row with different superscript letters differed significantly.

⁴ F-MPD₁ versus F-ESD ($p < 0.001$); F-MPD₂ versus F-ESD ($p < 0.001$); R-MPD versus F-ESD ($p = 0.002$); F-MPD₁ versus R-MPD ($p = 0.0102$).

⁵ F-MPD₁ versus F-ESD ($p < 0.001$); F-MPD₂ versus F-ESD ($p < 0.001$); R-MPD versus F-ESD ($p < 0.001$); F-MPD₁ versus R-MPD ($p = 0.0165$).

⁶ No significant differences.

⁷ F-MPD₁ versus F-ESD ($p < 0.001$); F-MPD₂ versus F-ESD ($p = 0.001$); R-MPD versus F-ESD ($p = 0.011$). Furthermore, the likelihood that an intrusion was a confabulation was significantly higher for White children than for Black children ($p = 0.032$).

Table 4

Number (and percent) of liking ratings¹ for each of the four groups² by item classification³ (match, omission, stretch, confabulation)

	Liking Ratings			Total
	Not at all	A little	A lot	
	----- n (percent) -----			
F-MPD ₁ (Study 1; 104 children)				
Matches	7 (2)	118 (31)	256 (67)	381
Omissions	80 (18)	139 (31)	228 (51)	447
Stretches	3 (7)	23 (51)	19 (42)	45
Confabulations	10 (4)	79 (34)	145 (62)	234
F-MPD ₂ (Study 2; 55 children)				
Matches	2 (1)	66 (33)	134 (66)	202
Omissions	29 (11)	97 (37)	133 (51)	259
Stretches	0 (0)	9 (45)	11 (55)	20
Confabulations	5 (5)	37 (37)	59 (58)	101
R-MPD (Study 2; 59 children)				
Matches	7 (3)	73 (32)	147 (65)	227
Omissions	43 (18)	74 (31)	122 (51)	239
Stretches	3 (15)	7 (35)	10 (50)	20
Confabulations	3 (4)	31 (40)	43 (56)	77
F-ESD (Study 3; 68 children)				
Matches	11 (3)	101 (27)	256 (70)	368
Omissions	30 (15)	58 (29)	109 (55)	197
Stretches	2 (6)	22 (67)	9 (27)	33
Confabulations	1 (2)	13 (27)	35 (71)	49
Total for all four groups ⁴				
Matches	27 (2)	358 (30)	793 (67)	1,178
Omissions	182 (16)	368 (32)	592 (52)	1,142
Stretches	8 (7)	61 (52)	49 (42)	118
Confabulations	19 (4)	160 (35)	282 (61)	461

¹ Refer to the Appendix for the specific prompts used to obtain liking ratings for each study.

² F-MPD interviews used **F**orward-order prompts (morning-to-evening) for meals/snacks eaten and were conducted in the **M**orning about the **P**revious **D**ay's intake; R-MPD interviews used **R**everse-order prompts (evening-to-morning) for meals/snacks eaten and were conducted in the **M**orning about the **P**revious **D**ay's intake; F-ESD interviews used **F**orward-order prompts for meals/snacks eaten and were conducted in the **E**vening about that **S**ame **D**ay's intake.

³ A *match* was an item that was observed eaten (in any amount more than none) and reported eaten (in any amount more than none) at a meal. An *omission* was an item that was observed eaten (in any amount more than none) but not reported eaten at a meal. An intrusion (an item reported eaten [in any amount more than none] that was not observed eaten) was further classified either as a *stretch* if it referred to an item that was observed on the tray for that meal, but for which the amount observed eaten was none, or as a *confabulation* if it referred to an item that was not observed on the tray for that meal.

⁴ A liking rating was not obtained for 12 matches, 88 omissions, 1 stretch, and 12 confabulations.