

On the use of wireless response systems in experimental psychology: Implications for the behavioral researcher

MOSES M. LANGLEY

Iowa State University of Science and Technology, Ames, Iowa

AND

ANNE M. CLEARY AND BOGDAN N. KOSTIC

Colorado State University, Fort Collins, Colorado

A method of data collection is presented that unites the efficiency of mass testing with the ease of instant electronic data collection that is typical of computer-based experiments run on individual participants. A wireless response system (WRS), originally designed as a teaching tool, is used to replicate three classic and robust effects from the memory literature (effects of false memory, levels of processing, and word frequency). It is shown that for these types of experimental designs, data can be collected more efficiently (in both time and effort) with the WRS method than through traditional mass- and individual-testing methods alone. The advantages and limitations of WRSs for use in mass electronic data collection are discussed.

Over the past decade or so, data collection in human experimental psychology has achieved a high level of sophistication; this sophistication has come about primarily through the use of computer technologies. Indeed, most modern-day experiments in human experimental psychology rely on computers for data collection. Before computers became a standard tool in behavioral experiments, many researchers relied on such materials as tape recorders for stimulus presentation (e.g., Anderson, 1974), reading stimuli aloud to participants (e.g., Morris, Bransford, & Franks, 1977), and visually presenting stimuli using words written on index cards (e.g., Slamecka & Graf, 1978). In many such cases, data were collected using paper and pencil techniques.

Today, the use of computers in experimentation allows for greater ease and efficiency in presenting, randomizing, and counterbalancing stimuli, as well as in recording, organizing, and storing data. Still, even with the use of computers in experimental research, many modern-day researchers often choose to use paper and pencil methods of data collection in order to collect data in groups of participants rather than individually. For example, in their now classic study of false memory, Roediger and McDermott (1995) read study list words aloud to a classroom full of students, and the students' memories for the words were tested using examination booklets. Other examples include the following: In her study of memory for meaning versus surface details of visual stimuli, Gernsbacher (1985) presented students in a small amphitheater with vi-

sual stimuli on a projection screen, and tested their memories for the stimuli using examination booklets. In their study of levels-of-processing effects on memory, Challis, Velichkovsky, and Craik (1996) distributed booklets of materials to groups of participants, which included encoding instructions and the task for the encoding phase of the experiment. Each participant also received a test booklet, used to test memory for the items used in the encoding phase of the experiment. Thapar and Greene (1994) studied levels-of-processing effects by presenting participants with written words on index cards while participants gave responses on numbered sheets of paper. More recent studies in the literature have tended to use Microsoft PowerPoint to present stimuli on a projection screen while testing with answer sheets (e.g., Carpenter & DeLosh, 2006; Voyer, Rodgers, & McCormick, 2004).

Whereas all of these types of experiments can easily be conducted from start to finish using computers, experiments that are completely computerized typically require running participants through experiments individually, which can result in delayed data acquisition (relative to running participants in groups). It is possible to conduct computerized experiments on groups of participants via networked workstations; however, there are drawbacks to this method relative to running participants in groups. For example, one would need access to a large number of networked computers in order to run an entire experiment in a single session. Furthermore, particularly in cases where a large sample size is required, even the use of networked

M. M. Langley, mlangley@iastate.edu

computers may require a large number of experimental sessions (whereas hundreds of participants can be run at once in an auditorium-like setting using a projection system). In addition, in order to synchronize the onset of an experiment among workstations and to centrally coordinate the data, one would need administrative access to the computers and the network, which may be difficult to gain depending on ownership of the lab. For these and other reasons, many present-day researchers choose collect data using paper and pencil methods in group settings via projection systems. The combination of projection systems and written tests in classroom settings allows for large amounts of data to be obtained in a single session, which is a highly efficient means of accruing data. However, the trade-off to having a more rapid data acquisition process is that time and energy must be spent in transcribing participants' responses from paper to digital/electronic form. Thus, when deciding on an experimental procedure, researchers must balance the costs and benefits of running participants individually using computers versus running participants in groups using classroom-like or auditorium-like settings.

In the present article, we present a means of integrating the advantages of running experiments in groups with the advantages of collecting data with computerized experiments. Here, we demonstrate that wireless response systems (WRSs), commonly used in teaching, can be used to collect data in group settings in digitized form, allowing for both the rapid data acquisition afforded by group-run experiments, and the rapid data extraction and analysis afforded by collecting data in digital format directly on a computer.

Wireless Response Systems

Also known as *audience/classroom response systems* or *classroom communication systems*, WRSs are packages of hardware and software that are generally intended to allow classroom instructors to receive instant feedback from their students. While there are various distributors of such systems, WRSs tend to work similarly. In general, these systems operate using wireless remote response pads that emit either an infrared (IR) or a radio frequency (RF) signal. These signals are detected by a small, portable receiver unit that connects directly to a laptop or desktop computer through the universal serial bus (USB) port. The software included in such packages enables a computer to process the remote signals such that the data for each response trial is recorded separately for each response pad (or for each student or participant). The software also generally allows for importation of the resulting data file into a spreadsheet file, such as a Microsoft Excel file, for data sorting, analysis and storage. The number of remote key pads included with each system varies, but the capacity of most systems to simultaneously process incoming data ranges between 70 and 270 signals. With the ability of WRSs to process such a large number of responses simultaneously, it is clear that such systems may have the potential to provide a very efficient means of data collection in behavioral research.

Moreover, because many academic institutions are beginning to integrate WRSs into their classrooms for teach-

ing purposes, WRSs are becoming increasingly available to members of academic institutions. The growing use of WRSs at academic institutions makes the application of this teaching tool to behavioral research particularly feasible, as many researchers may currently have access or may soon have access to such systems at little or no cost.

It is likely that the wireless response method of data collection can be applied to many forms of psychological research, including nonexperimental methods (e.g., survey research, norming of stimuli) as well as experimental methods. In fact, one study (Bunz, 2005) has recently assessed the reliability of a WRS in the collection of nonexperimental data. In her comparison of survey data collected through a traditional method (scantron sheets) versus a WRS method, Bunz reported that the former method produced data that was no less reliable than data produced from the latter method. Moreover, Bunz also reported (1) few disadvantages of the new WRS method and (2) that participants found the WRS more fun than and of equivalent ease to the scantron method.

To our knowledge, WRSs have not been used in published research in experimental psychology, and, consequently, the efficacy of such a technique in experimental research is unknown. Previous efforts to incorporate WRSs in the collection of survey data have proven successful and advantageous (Bunz, 2005), a finding that suggests similar advantages may also be afforded to the collection of experimental data. Thus, the goal of the present article was to explore this method of data collection by comparing data collected in this manner with data collected through more traditional and established methods (e.g., individual computer testing and group testing via paper and pencil techniques). Toward this end, we examined whether classic findings from the literature could be replicated using the WRS.

For experimental psychologists, this method may be particularly useful to those employing list-learning paradigms in studies of human memory. One reason is that, unlike some areas of experimental psychology, list learning methods often involve stimulus presentation durations that are one to several seconds in length (as opposed to rapid flashes that occur on the order of milliseconds, which cannot be done using classroom projection systems). Another reason is that list learning methods typically use accuracy measures (as opposed to response time measures, which cannot be carried out with remote keypads). For these reasons, we attempted to replicate three classic memory phenomena using a WRS.

Demonstrating that known findings can be replicated using the WRS method is important because it is unknown what effects, if any, mass testing with wireless remote response pads might have on experimental results. In considering the literature on group behavior, it is conceivable that mass-testing with wireless response systems might have some effect on participants' performance. The classic conformity experiments in social psychology suggest that people may at times have a tendency to behave in a manner accordant with the group (e.g., Sherif, 1936); thus, conformity may play a role in people's performance, particularly in cases where feedback about the group's

performance appears automatically on the screen after the whole group has responded to an item (as occurs with the Classroom Performance System when integrated with PowerPoint, described below). Because such conformity is not likely an issue in situations where participants are run individually on computers, or in group settings where paper and pencil tests are used, it is important to determine if already known experimental findings can be replicated using a wireless remote keypad response system.

Microsoft Office PowerPoint

The WRS used in the present set of studies was the Classroom Performance System (CPS). As may be true for other WRSs, the CPS was designed to interface with Microsoft Office PowerPoint, which we have used to present stimuli during the encoding phase of Experiment 3. An increasing number of studies in experimental psychology have been using PowerPoint for the purposes of stimulus presentation in group settings (e.g., Carpenter & DeLosh, 2006; Voyer, Rodgers, & McCormick, 2004). Thus, for those researchers already relying on PP for stimulus presentation, the ability of some WRSs to interface with PP may be an additional benefit of the WRS method.

EXPERIMENT 1A

Experiment 1A was an attempt to replicate the classic false memory effect reported by Roediger and McDermott (1995) using a WRS in a group setting. Roediger and McDermott had participants study lists of words (e.g., *bed, rest, awake, tired*) that were all semantically related to a particular "critical" nonpresented word (e.g., *sleep*). On the subsequent memory test, participants were more likely to respond "old" to critical nonstudied words (e.g., *sleep*) than to nonstudied words that did not relate to earlier presented lists (e.g., *shovel*). The primary difference in methodology between our study and that of Roediger and McDermott was that our memory test was administered using a projection system and wireless response pads rather than with paper and pencil.

Method

Participants. In Experiment 1A, 31 undergraduate students participated to fulfill a requirement for their introductory psychology course. In Experiment 1B, 10 undergraduates participated.

Materials. Ten 15-word lists were chosen from the appendix of Roediger and McDermott's (1995) manuscript. The words were digitally recorded in a female voice using a digital microphone and saved in a .wav format for use in the experiment. At study, E-Prime Version 1.1 software was used to present five 15-item lists auditorily through a classroom audio system (which contained ceiling speakers distributed throughout the classroom). At test, 15 test cues (five old words, five critical new words, and five noncritical new words) were presented in a random order visually via a classroom projection screen using the CPS software. All auditory and visual presentations were controlled and executed from a Dell Inspiron 8100 laptop computer, and the data were collected through this computer as well (via the IR receiver unit connected to the laptop).

Procedure. The present experiment was conducted in a medium-sized classroom, with seating for approximately 90 students. At the start of the experiment, each participant was assigned a numbered remote keypad. Participants were then shown a short instructional

audiovisual presentation (presented using E-Prime Version 1.1 software) explaining that they were participating in a memory experiment and that their responses on the memory test would be collected through the use of the wireless keypads. Following the instructional presentation, participants were presented with the study list. The pre-recorded study list was presented auditorily (via E-Prime Version 1.1) through a set of speakers that were part of the classroom's audiovisual system. The items within each 15-word list were associates of a single "critical word" that was not actually present in the list.

Following the auditory study-list presentation, the visual memory test was presented on a large projection screen via digital video projector. At this time participants were informed that their memory would be tested for items from the study list. The test list consisted of one studied word from each of the five study lists (totaling five old items), the five critical words corresponding to each of the five study lists, and five critical words corresponding to word associate lists not presented at study (totaling five new items). Test-list words were presented one at a time on the screen with instructions asking the participants to judge whether the item had been present in the preceding auditory study list. To respond, the participants were instructed to press the "A" key for a "yes" response and the "B" key for a "no" response. The duration between each test word was group paced; the experimenter waited until each participant had responded before proceeding to the next test item. As the experimenter attempted to move from one test item to the next, a graph appeared on the screen showing the number of responses in each category (this is an automatic and incompressible feature of the CPS software).

Data extraction and analysis. Data for all 31 participants were recorded to a single Microsoft Excel compatible file by the software associated with CPS. By default, this particular system's software organized response data by participant (rows) and by test cue (columns), with participants ordered sequentially by the number of their keypad, and with test cues presented linearly in the same sequential order as presented to participants.

When one is finished with a CPS session, the corresponding data file can be opened by clicking on the data file within the Reports file of the CPS software system. Clicking on the data file will open a list of options for viewing the data. For example, one can view item-by-item percentages, graphs of overall responses, and other descriptive statistics. For the present experiment, the data were opened using the option called "Response Data Export," which exports the data to an Excel CSV (comma separated values) file. From that file, the raw data were copied and pasted into a previously created Excel file containing formulas that would be applied to the data upon pasting. This Excel file was a template that had been created prior to the demonstration to allow for rapid analysis of the data. The Excel formulas first converted the "A" and "B" responses to 1 and 0 sec, respectively. The converted data were then sorted according to test item type (studied items, critical nonstudied items, and noncritical nonstudied items). The sorted data were then pasted into a separate worksheet where existing formulas computed the proportion of "old" responses given to each item-type for each student. From these data, a separate set of Excel formulas were used to compute the overall average proportion of "old" responses for each test-item category (old, critical, and new).

Results and Discussion

The mean proportions of items called "old" in each category are shown in Figure 1. As can be seen in Figure 1, the mean proportion of old responses given to studied words ($M = .83$, $SD = .19$) was greater than the proportion given to critical words ($M = .75$, $SD = .21$), which was greater than that given to new words ($M = .25$, $SD = .23$) [$F(2,60) = 3.40$, $MS_e = .05$, $p < .001$]. The difference between the proportion of old responses to studied words and critical words was unreliable [$t(30) = 1.72$, $SE = .05$, $p = .10$]. Most important for the present

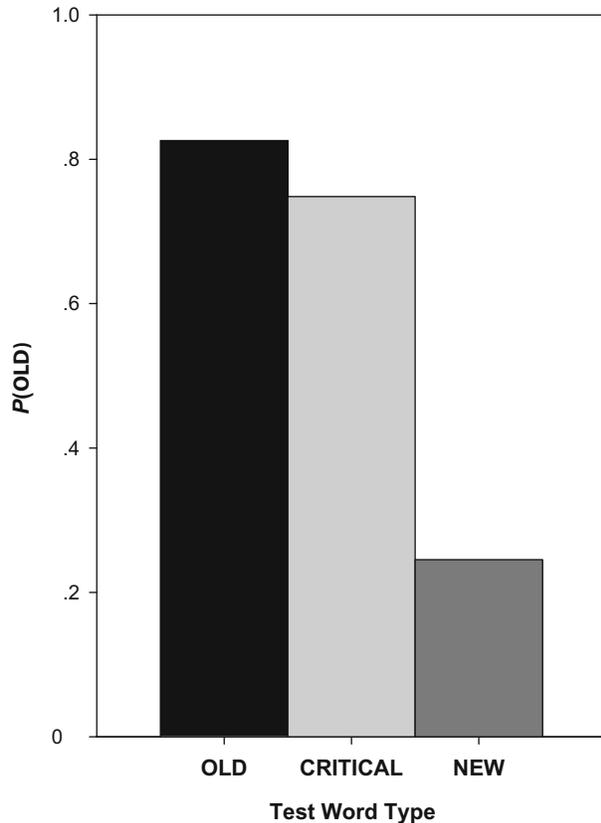


Figure 1. Replication in Experiment 1A of the false memory effects shown by Roediger and McDermott (1995). On the y-axis is the mean proportion of “old” (“yes”) responses given to test items. On the x-axis are the three types of words presented at test.

purposes, however, was the comparison between critical items and new items. The difference between the proportion of old responses given to these two types of test cues was significant [$t(30) = 8.39$, $SE = .06$, $p < .001$] (Cohen’s $d = 1.50$), indicating that false memory for critical items occurred.

The replication of this classic false memory effect with the WRS suggests that data obtained through this method may generally be comparable to large group data collected via paper and pencil as well as to data collected individually using computers. Thus, it appears that WRS may in fact be viable method of running large group experiments in behavioral research.

EXPERIMENT 1B

As with all memory experiments run on large groups in a single session, counterbalancing the stimuli across conditions can be an issue with data collection using WRS methods. For example, in addition to receiving the same items in the same order, participants in group-run experiments receive the same stimuli in the same conditions. This can result in item effects if counterbalancing of the stimuli across conditions is not performed. The purpose of Experiment 1B therefore was to demonstrate how coun-

terbalancing can be used to remedy this inherent issue in large group experiment sessions.

In the procedure of Experiment 1A, 5 of 10 potential word lists were presented auditorily at encoding. At test, five studied words were presented, five critical words, and five new words (which were the critical words corresponding to non-presented lists from Roediger & McDermott, 1995). Because the stimuli were not counterbalanced across conditions, in theory, the results could reflect greater inherent familiarity of the particular critical items used, rather than their relation to studied items. Therefore, a second variation of Experiment 1A was carried out in which the stimuli were counterbalanced across conditions.

Method

For the purpose of counterbalancing in Experiment 1B, two sessions of 5 participants each were run. The first group of participants received one set of five lists at encoding and the second group received the other set of five lists (those for which the critical items had served as new items for the first group). Thus, at test, both groups received five studied words and five critical words corresponding to their respective studied lists. The five new words for a given group were the five critical items used at test for the alternative group.

Results and Discussion

The mean proportion of old responses given to studied words ($M = .74$, $SD = .25$) was greater than the proportion of old response given to critical words ($M = .64$, $SD = .21$), which was greater than the proportion given to new words ($M = .22$, $SD = .18$); this difference across word types was significant [$F(1,9) = 42.25$, $MS_e = .03$, $p < .001$]. As in Experiment 1A, the difference between the average proportion of old responses given to critical and new words was significant ($M = .42$, $SD = .017$) [$t(9) = 7.58$, $SE = .06$, $p < .001$] (Cohen’s $d = 2.34$).

Thus far, we have demonstrated that the use of WRS in behavioral research is methodologically feasible, highly efficient, and produces comparable results to those found in the literature when replications are attempted. Moreover, the counterbalancing issue inherent in group designs can be overcome by running at least two groups and counterbalancing the stimuli across conditions for these groups.

EXPERIMENT 2

Having demonstrated the false memory phenomenon using the WRS method, Experiment 2 served to replicate another classic memory finding using the WRS method. Specifically, Experiment 2 served to replicate the word frequency mirror effect in recognition memory (Glanzer & Adams, 1990). The word frequency mirror effect is the finding that low-frequency words tend to show higher hit rates and lower false alarm rates than high-frequency words on tests of recognition memory for list items. Another purpose of Experiment 2 was to compare the effect sizes obtained using the WRS method with those obtained using more traditional methods of data collection. Whereas a search of the literature revealed no effect size information for false memory effects (which were exam-

ined in Experiments 1A and 1B of the present study), such information is available in the literature for reports of the word frequency mirror effect. Thus, such an effect size comparison could be made in Experiment 2. Comparing effect sizes should allow for an assessment of whether the WRS method might lead to effects that are comparable in magnitude to those found using other, more commonly used methods.

Method

Participants. Fifty-four undergraduate students participated to fulfill a requirement for their introductory psychology course.

Materials. The equipment used for Experiment 2 was the same as that used in Experiments 1A and 1B. Forty words taken from the Kučera and Francis (1967) norms were used in this experiment; 20 of these words were high frequency ($M = 96.20$) and 20 were low frequency ($M = 1.20$). These stimuli were counterbalanced across four experimental sessions. During encoding, 20 words were presented (10 high frequency, 10 low frequency). During memory testing, all 40 words were presented.

Procedure. The procedure surrounding the use of the WRSs themselves was the same as described in Experiment 1, with the following exceptions. Participants viewed a series of high- and low-frequency words during encoding. These words were presented individually on a projection screen in the front of the classroom at a rate of 2 sec per word (with an interstimulus interval of 1 sec) using E-Prime Version 1.1. Participants were instructed to remember the words for a later memory test. As in Experiments 1A and 1B, participants were given a visual yes–no recognition test using the CPS.

Results and Discussion

The results of Experiment 2 are presented in Figure 2. For the purposes of assessing the magnitude of the effects obtained with the WRS relative to effects obtained through other data collection methods, we report Cohen's d (Cohen, 1988). In general, the lowest values of d imply small effect sizes ($d < 0.39$), moderate values imply medium effect sizes ($0.4 < d < 0.69$), and high values imply large effect sizes ($d > 0.7$). As can be seen here, the standard word frequency mirror effect was shown. The average proportion of "yes" responses given to studied high-frequency words (i.e., the hit rate) was .78 ($SD = .18$) whereas the average proportion of "yes" responses given to studied low-frequency words was .84 ($SD = .14$) [$t(53) = 2.86$, $SE = .02$, $p < .01$] ($d = 0.40$). Additionally, a higher false alarm rate was shown for high-frequency words ($M = .17$, $SD = .20$) than for low-frequency words ($M = .11$, $SD = .14$) [$t(53) = 3.30$, $SE = .02$, $p < .01$] ($d = 0.45$).

The individual components of the word frequency mirror effect replicate the data pattern indicative of the word frequency mirror effect as a whole, both in terms of the direction of means reported in the literature (e.g., higher hit rates for low-frequency words) and the magnitude of the effect size reported in the literature. For example, Wilhite (1981) examined word frequency effects using a group testing format and a sample size comparable to ours ($n = 40$). However, Wilhite administered stimulus presentation via note card and memory testing via paper and pencil. Despite the differences in the encoding and testing conditions between his study and ours, Wilhite found similarly sized effects for the hit rate advantage of low-frequency

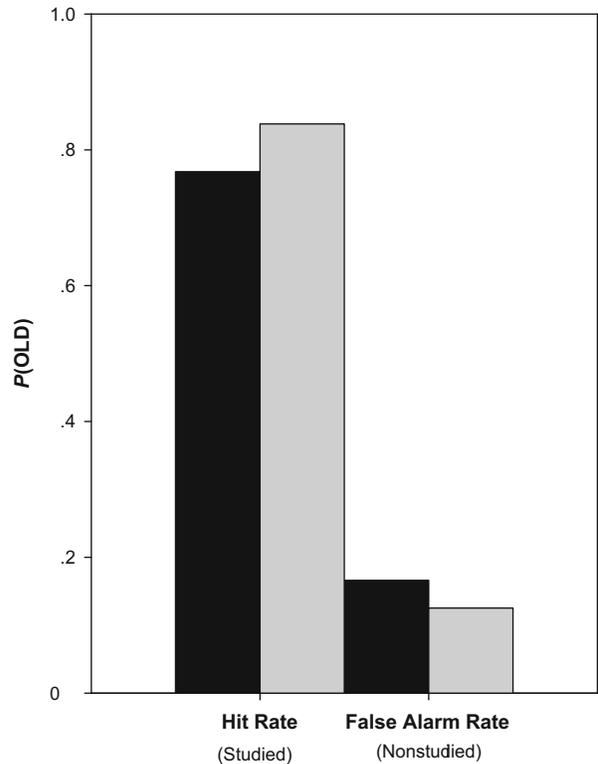


Figure 2. Replication in Experiment 2 of the robust word frequency effect (e.g., Glanzer & Adams, 1990). On the y-axis is the mean proportion of "old" ("yes") responses given to test items. The x-axis delineates the proportion of "old" response given to studied and nonstudied words. The black bars represent recognition performance on high-frequency words and the gray bars represent recognition performance on low-frequency words.

words ($d = 0.45$) and the higher false alarm rate for high-frequency words ($d = 0.29$). In our review of the literature, medium to small values of Cohen's d are typical effect sizes for each component of the word frequency mirror effect (e.g., Glanzer & Adams, 1990); as such, the medium sized effects obtained with the WRS in the present experiment are consistent with the effects reported in the literature. Moreover, the effects reported here are consistent with effects obtained in other forms of group-run experiments (Wilhite, 1981).

As to the word frequency mirror effect as a whole, the effect itself is defined as both higher hits and lower false alarms for low-frequency words than for high-frequency words. Consequently, the difference between the hits and false alarms should be greater for low-frequency words. To confirm the integrity of the mirror pattern within the hits and false alarms, the discrimination measure P_r (hits – false alarms) was calculated for high-frequency ($M = .60$, $SD = .27$) and low-frequency ($M = .78$, $SD = .22$) words. A paired samples t test confirmed a larger P_r value for low-frequency words [$t(53) = 4.47$, $SE = .03$, $p < .001$] ($d = 0.62$). The medium effect size of the full mirror pattern is also typical of reports in the literature. For example, Clark (1992) also examined word frequency effects but did so by examining participants individually in comput-

erized experiments. Clark, using a sample size similar to ours ($n = 48$), also observed “medium” word frequency mirror effects ($d = 0.46$). Thus, the full pattern of results shows a standard word frequency mirror effect.

The conformity of general effect size (d) in the present study (at the level of the mirror effect’s individual components and at the level of the mirror effect as a whole) with various similar studies in the literature complements the present experiment’s replication of the effects themselves. It would appear, therefore, that data collected with a WRS under the present conditions is comparable on many levels with data collected through more traditional means. This finding suggests that the WRS may be a reliable and valuable data collection tool.

EXPERIMENT 3

Experiment 3 sought to further examine the feasibility of the WRS method for mass data collection in memory research by attempting to replicate the classic levels-of-processing effect (Craik & Tulving, 1975). This would demonstrate the method’s use in encoding manipulations that require responses, as well as in memory testing itself, in group settings.

Method

Participants. Eighteen undergraduate students participated to fulfill a requirement for their introductory psychology course.

Materials. The equipment and software was the same as in Experiments 1A, 1B, and 2. Thirty words were used as stimuli in Experiment 3. These words were selected from the Kučera and Francis (1967) norms. Twenty words were presented during the encoding phase, 10 of which were encoded under shallow encoding instructions and the remaining 10 were encoded under deep encoding instructions. The test phase consisted of all 20 studied words and 10 new words and was carried out in the same manner as in Experiments 1 and 2.

Procedure. During the encoding phase, participants were presented with 20 words on the projection screen. For half of these words, participants were instructed to count the number of vowels (shallow encoding). They reported the number of vowels using wireless remote keypads (A = 1, B = 2, C = 3, D = 4, E = 5). For the remaining items on the study list, participants rated the pleasantness of the words (deep encoding) on a scale of 1–5 (1 = *very unpleasant*, 5 = *very pleasant*; A = 1, B = 2, C = 3, D = 4, E = 5). The stimuli and their corresponding encoding instructions were present in random order and were presented using Microsoft Office PowerPoint. During the retrieval phase, participants completed a yes–no recognition task for thirty words (20 studied, 10 unstudied in a random order) and were asked to press “A” for “yes” and “B” for “no.”

Results and Discussion

Experiment 3 replicated the standard levels of processing effect by showing a greater proportion of yes responses for deeply encoded items ($M = .93$, $SE = .01$) than for shallowly encoded items ($M = .35$, $SE = .04$) and a greater proportion of yes responses for shallowly encoded items than for new items ($M = .04$, $SE = .02$) [$F(1,17) = 1,110.20$, $MS_e = .006$, $p < .01$] (see Figure 3). The effect size for the contrast between deeply encoded and shallowly encoded items was very large [$t(17) = 12.53$, $p < .01$; $d = 2.90$], as has been found in group-run studies administered via paper and pencil (e.g., Gardiner,

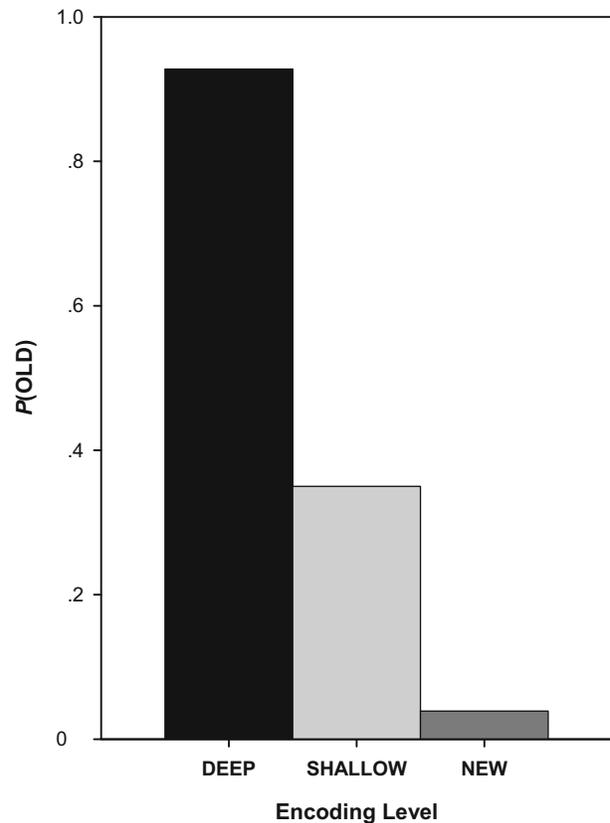


Figure 3. Replication in Experiment 3 of the levels-of-processing-type effects shown by Craik and Lockhart (1972). On the y-axis is the mean proportion of “old” (“yes”) responses given to test items. On the x-axis are the three types of test words presented at test.

Java, & Richardson-Klavehn, 1996; $d = 2.52$) as well as individually run studies administered via computer (e.g., Bodner & Lindsay, 2003; $d = 2.14$). The results of Experiment 3 are consistent with those of Experiment 2 in that the particular effects were not only replicated with the WRS but were also found to be of the same magnitude as effects obtained through traditional means that differ considerably from the present methodology.

GENERAL DISCUSSION

The aim of the present study was to assess the utility of wireless response systems for collecting data in experimental behavioral research. In four experiments, we replicated several classic findings from the literature using a WRS. These findings were (1) the false memory phenomenon (Roediger & McDermott, 1995), (2) the word frequency mirror effect in recognition memory (e.g., Glanzer & Adams, 1990), and (3) the levels-of-processing effect (Craik & Tulving, 1975). Furthermore, the effect sizes reported in the literature (for both group-run and individually run experiments) were consistent with the effect sizes obtained with the WRS. The fact that these experiments could all be replicated using the WRS method suggests that the WRS may be a viable means of mass data collection in memory research. Furthermore, use

of the WRS method should allow researchers to obtain effects that are comparable in magnitude to those found using other methods (e.g., paper and pencil methods of group testing and individually run computerized experiments).

With the arrival of the WRS in university classroom settings, researchers may no longer face a time trade-off to running behavioral experiments in large groups at a time. Specifically, the rapid data acquisition that comes with running participants in groups (relative to running participants individually on computers) may no longer involve a lengthy process of transcribing participants' responses from paper to digital/electronic form. The WRS method allows for a large quantity of data to be collected in a short period (as occurs in group-run experiments) while at the same time, doing so digitally, and in a format that can be opened immediately afterward for sorting and analysis in Microsoft Excel or an equivalent spreadsheet document.

The WRS method does have its limitations and disadvantages; however, many of these limitations and disadvantages are the same as those present in any group-run memory experiment. For example, the issue of counterbalancing the stimuli across conditions is one that needs to be dealt with in any group-run memory experiment, and this can be done in the method presented in Experiment 1B. Also, participants in a group receive the stimuli in the same order, which will work to inflate any particular item order effects. Another limitation that the WRS has in common with paper and pencil group testing methods is that reaction times cannot easily be recorded. Therefore, though future variations of WRSs may contain this capability, the method is not currently useful for reaction time studies.

As is present in any group-run memory experiment, there is also the issue of participants asking questions during the experiment. In many experiments, participants are allowed and even encouraged to ask questions if they do not understand the instructions. In a mass-testing environment, one participant asking a question during the encoding or the testing phase could potentially be disruptive to the experiment. There is also the threat of participants communicating with one another. Studies on group collaboration and memory have shown that in many instances, group collaboration can actually impair recall compared to individual performance (e.g., Weldon & Bellinger, 1997). Obviously, participants should be asked not to speak to one another during the experiment, but one cannot always control people who choose to chat or whisper. Therefore, this could potentially be an issue when trying to run large groups in a WRS setting.

One limitation that is unique to the WRS method (that is not present in paper and pencil methods of group testing) is that participants can only provide single-key responses using the remote keypads. This means that participants cannot type whole words, as they might on a recall test or for an identification task. Therefore, though future variations of the WRS method may include the ability to type string responses, at the present time, the WRS method is not useful for studies that require multiple-key responses (as when a participant must type in a word).

Another issue (which may be unique to CPS) is that, when used alone or in conjunction with Microsoft Power-

Point, the CPS software automatically provides group feedback in between test items. That is, as the experiment is shifting from one slide to the next, a graph or chart automatically appears displaying the percentage of responses in each category. This was the case in the experiments reported here; however, as can be seen in the data themselves, this did not impact the findings. We still replicated the basic findings that we were attempting to replicate. However, to avoid presenting feedback to the group with the CPS, the test items can be presented directly through the CPS software rather than through PowerPoint. With this presentation method, it is easy to prevent displaying feedback to participants altogether.

Some might argue that our finding that the WRS method can be used to replicate known findings from list-learning experiments is unsurprising. However, without an existing demonstration that the method itself does not influence one's findings, it may be difficult for a researcher attempting to use this method to produce novel empirical findings to make the case that it is a viable method of data collection. It is our hope that the present article will help to demonstrate that the WRS is a viable method of collecting data in memory research. The fact that this new method can be used to replicate known findings from the memory literature—findings that have been obtained in group settings as well as individuals on computers—suggests that it is a viable means of carrying out original research projects.

Admittedly, the present investigation examined use of WRS under a somewhat narrow range of experimental procedures, but it is easy to see how this method might be used with other measures of memory and in other areas of psychological research. For example, WRSs can be used to give ratings responses to test items (such as confidence ratings), and thus can be used to construct ROC curves. They can also be used in forced-choice testing situations. They can even be used for testing memory for location information presented visually on the screen. For example, participants might be presented with a 5×5 grid on the screen. To test memory for location, the experiment might first present the possible columns (A, B, C, D, or E) and then following participants' responses to that screen, present the possible rows (A, B, C, D, or E) so that participants can indicate location on the grid.

As far as other areas of research are concerned, the present method is likely conducive to studies of eyewitness memory and perpetrator identification. Such a study might involve participants viewing a video of a crime being committed. Subsequently, participants' memories are tested for information in the video, or participants are asked to select the perpetrator from a series of mug shots. The present method may also be conducive to studies of stereotype and bias, in which large numbers of participants are asked to indicate their beliefs about or impressions of stigmatized groups by selecting a descriptor from a presented list. It may also be useful for some types of survey research to save the step of transcribing paper-based data into a digital format.

The possible applications of WRS method to the various procedures used in experimental psychology are of

course numerous, and it is likely that with future variations of WRSs, future researchers will devise ways to use this methodology more creatively and effectively than what has been described here. However, what is presented here is intended to support the validity of this method for use in data collection in experimental psychology. For whom this method is applicable, we advocate this method's use data in the collection, as a means of efficiency in time and labor.

AUTHOR NOTE

This study was part of a research program supported by National Science Foundation Grant 0349088. The authors thank Neal R. Iverson. Thanks are also due Jeremy Olhart, Cynthia Brownlee, and Michelle McDonald for their assistance with data collection. None of the authors has a commercial or financial interest in the software or hardware discussed in this article. Correspondence concerning this article should be addressed to M. M. Langley, Department of Psychology, W 112 Lagomarcino Hall, Iowa State University of Science and Technology, Ames, IA 50011-3180 (e-mail: mlangley@iastate.edu).

REFERENCES

- ANDERSON, J. R. (1974). Verbatim and propositional representation of sentences in immediate and long-term memory. *Journal of Verbal Learning & Verbal Behavior*, **13**, 149-162.
- BODNER, G. E., & LINDSAY, D. S. (2003). Remembering and knowing in context. *Journal of Memory & Language*, **48**, 563-580.
- BUNZ, U. (2005). Using scantron versus an audience response system for survey research: Does methodology matter when measuring computer-mediated communication competence? *Computers in Human Behavior*, **21**, 343-359.
- CARPENTER, S. K., & DELOSH, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition*, **34**, 268-276.
- CHALLIS, B. H., VELICHKOVSKY, B. M., & CRAIK, F. I. M. (1996). Levels-of-processing effects on a variety of memory tasks: New findings and theoretical implications. *Consciousness & Cognition*, **5**, 142-164.
- CLARK, S. E. (1992). Word frequency effects in associative and item recognition. *Memory & Cognition*, **20**, 231-243.
- COHEN, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- CRAIK, F. I. M., & TULVING, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, **104**, 268-294.
- GARDINER, J. M., JAVA, R. I., & RICHARDSON-KLAVEHN, A. (1996). How level of processing really influences awareness in recognition memory. *Canadian Journal of Experimental Psychology*, **50**, 114-122.
- GERNSBACHER, M. A. (1985). Surface information loss in comprehension. *Cognitive Psychology*, **17**, 324-363.
- GLANZER, M., & ADAMS, J. K. (1990). The mirror effect in recognition memory: Data and theory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 5-16.
- KUČERA, H., & FRANCIS, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- MORRIS, C. D., BRANSFORD, J. D., & FRANKS, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning & Verbal Behavior*, **16**, 519-533.
- ROEDIGER, H. L., III, & McDERMOTT, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 803-814.
- SHERIF, M. (1936). *The psychology of social norms*. New York: Harper.
- SLAMECKA, N. J., & GRAF, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning & Memory*, **4**, 592-604.
- THAPAR, A., & GREENE, R. L. (1994). Effects of level of processing on implicit and explicit tests. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 671-679.
- VOYER, D., RODGERS, M. A., & MCCORMICK, P. A. (2004). Timing conditions and the magnitude of gender differences on the metal rotation test. *Memory & Cognition*, **32**, 72-82.
- WELDON, M., & BELLINGER, K. D. (1997). Collective memory: Collaborative and individual processes in remembering. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **23**, 1160-1175.
- WILHITE, S. C. (1981). Word-frequency cueing effects: Recognition and encoding interference factors. *American Journal of Psychology*, **94**, 323-337.

(Manuscript received October 22, 2006;
revision accepted for publication January 25, 2007.)