A Comparison of Two Procedures for Assessing Preference in a Classroom Setting

Keith C. Radley¹ · Evan H. Dart¹ · Allison A. Battaglia¹ · W. Blake Ford¹

Published online: 18 April 2018
© Association for Behavior Analysis International 2018

Abstract

The purpose of this study was to compare a method of assessing preference within a large group format to individual preference assessments. Individual preference assessments were conducted by presenting an array of four edible stimuli to a participant and allowing the participant to select a preferred stimulus, with stimuli removed from the array based on selection criteria. Group preference assessments were conducted in a classroom of 19 students, with all students responding simultaneously to a prompt to identify a preferred stimulus using Plickers—unique Quick Response code cards that are read by an accompanying smartphone app. During the group procedure, stimuli in the array were restricted on the individual participant level. Results indicated that the group procedure was a valid and rapid method of assessing preference within a group of individuals. Although additional research is required, practitioners and researchers may consider use of Plickers as a promising means of evaluating preference within a group setting.

Keywords Individual reference assessment · Group preference assessment

Group contingencies are frequently utilized to modify student behavior at a school-wide and class-wide level. For example, a teacher may use the Good Behavior Game, in which students may earn points that allow access to a reinforcer (Barrish, Wolf, & Saunders, 1969). At the school level, students may earn tokens for exhibiting desirable behavior, which may then be exchanged later for backup reinforcers (Bradshaw, Reinke, Browns, Blevins, & Leaf, 2008). School-based group contingencies are time and resource efficient, as a single reward is typically delivered to all members of a group (Gresham & Gresham, 1982); however, the effectiveness of such programs for any individual student or group of students is dependent upon idiosyncratic preferences for stimuli delivered as contingent rewards (Wehman, 1976). Put another way, access to edibles, tangibles, or activities available through programs like the Good Behavior Game may function as a reinforcer for some students' behavior, whereas such access may not function as a reinforcer for other students' behavior—reducing its effectiveness in modifying group behavior. Although

not explicitly required by interventions employing class-wide group contingencies, teachers often select a single stimulus or activity that will be used as reinforcement (e.g., Dart et al., 2016) or randomly select an item from a list of stimuli (e.g., Hawkins, Haydon, Denune, Larkin, & Fite, 2015). The drawback to this method of stimulus selection is that it does not consider student preference.

Group contingencies have a large body of literature supporting their effectiveness in schools without consideration of student preference (e.g., Maggin, Johnson, Chafouleas, Ruberto, & Berggren, 2012); however, behavior change is rarely, if ever, assessed at the level of the individual student and estimates of group behavior are used instead. Thus, increases in the academically engaged behavior of a group of students under the effect of a group contingency may only be reflective of behavior change in a small group of students. Additionally, many interdependent group contingencies, which are the most commonly implemented within a school setting (Maggin et al., 2012), do not require all members of the group to modify their behavior in order for the contingency to be satisfied. In both cases, it is impossible to determine whether all students exhibit a behavior change without individual behavior assessments, which can be extremely time consuming. So, to increase the likelihood that interventions utilizing group contingencies are effective for

✉ Keith C. Radley
keith.radley@usm.edu

¹ Department of Psychology, University of Southern Mississippi, 118 College Drive # 5025, Hattiesburg, MS 39406-001, USA

all students, individual student preference for reinforcers may be considered to promote uniform effectiveness. Preference of individuals is a critical element to consider when developing and implementing behavior change procedures due to its relation with reinforcer efficacy. Although delivery of highly preferred stimuli as consequences of behavior does not guarantee the efficacy of an intervention, researchers have concluded that preference of stimuli is a strong indicator of reinforcement efficacy (Piazza, Roane, & Karsten, 2011). Given the potential of stimulus preference to impact an individual's response to intervention, methods have been devised to systematically evaluate preference for stimuli (Roane, Vollmer, Ringdahl, & Marcus, 1998).

Preference assessments describe a class of methodologies used to identify potentially reinforcing stimuli, activities, or contexts (Cooper et al., 2007). Preference assessment procedures are typically conducted with an individual participant and involve the presentation of stimuli and the recording of stimulus selection across several trials. The results of repeated presentation and recording of selections are then commonly used to create a preference hierarchy, which may be used to estimate the reinforcing value of a stimulus. Although preference assessments are frequently used to inform preference within individual reinforcement contingencies (Kodak, Fisher, Paden, & Dickes, 2013; Krentz, Miltenberger, & Valbuena, 2016), such procedures are rarely described in the literature for identification of potential reinforcers for use in a group contingency (e.g., Good Behavior Game). When preference is assessed within a group setting, it is often done through use of reward surveys with limited detail regarding selection and delivery of preferred stimuli (e.g., Hirsch, Healy, Judge, & Lloyd, 2016; Lannie & McCurdy, 2007).

The limited use of preference assessment procedures within group contingencies is likely due to the prohibitive amount of time that would be necessary to conduct individual preference assessments with a large group. For example, Layer, Hanley, Heal, and Tiger (2008) estimate that conducting preference assessments using previously established procedures (Heal & Hanley, 2007) with each student within a 20-student classroom would take more than 140 h. In order to overcome this obstacle, Layer et al. (2008) described a group preference assessment in which three children simultaneously identified a preferred stimulus using colored cards, which were placed in a box. A lottery-type system was then used, as one colored card was drawn from the box and the corresponding stimulus was delivered to each of the three children in the group. With the use of the lottery system, the group preference assessment described by Layer and colleagues deviated from individual preference assessments in which selected stimuli are delivered on a continuous schedule. Instead, the delivery of a selected stimulus was probabilistic—with the probability of a stimulus being delivered dependent upon the number of stimuli available for

selection and the selections of other children in the three-student arrangements.

Although the purpose of the Layer et al.'s (2008) study was to assess children's preference for shared contexts, the results indicated that the group preference assessment was a more efficient means of assessing preference of all members of a group. Additionally, probabilistic delivery of stimuli as reinforcement potentially addresses the issue of marginalizing students who indicated a preference for stimuli not selected by the majority by ensuring that there is at least a small chance their preferred stimulus is delivered. Not only did group preference assessment procedures require fewer assessment sessions (14 individual sessions compared to 5 group sessions), the researchers found preferences to be more highly discriminated (i.e., individuals demonstrated low variability in stimulus selection within a preference assessment) within the group preference assessment procedure; however, because group preference assessments were only conducted with three students at a time, this procedure required multiple group preference assessments to be conducted before a hierarchy could be developed for all students. Additionally, it is not clear how group size would impact stimulus selection when reinforcer delivery is probabilistic in the group arrangement. That is, in a group of three students, there is a basal probability of 33% that an individual will receive their preferred stimulus if they are the only one to select that stimulus. Similarly, in a group of 5, 10, or 20 students, the basal probability of an individual receiving their preferred stimulus decreases (e.g., 20, 10, 5%, respectively) if they are the only one selecting that stimulus. These probabilities also vary as a function of array size. That is, as more stimuli are added to an array, it is more likely that all stimuli receive fewer selections. It is not uncommon for classrooms to contain 20 or more students, and it is necessary to determine if the results of group preference assessments in larger groups still correspond to the results of individual preference assessments as they did in Layer et al. (2008).

Purpose

The purpose of this study was to evaluate a method of conducting a group preference assessment in a large group (i.e., whole class) in comparison with individual preference assessments using procedures adapted from Layer et al. (2008). It was hypothesized that the group preference assessment procedure would result in lower indifference scores, as well as preference assessments that corresponded with individual preference assessments in terms of first item selected and preference hierarchies. Additionally, this study sought to evaluate the feasibility of group preference assessment utilizing Plickers, a smartphone application that automatically records responses to each question or trial, as a means of

collecting selection data during group preference assessment procedures. It was hypothesized that group preference assessments incorporating Plickers would result in a lower amount of time to identify preferences of all group members than individual preference assessments.

Methods

Participants and Setting

Participants included 19 children in a 7th grade classroom in the Southeast region of the USA. All participants were between the ages of 12 and 14 years. The classroom was referred for services due to high levels of disruptive behavior. Prior to data collection, consent and assent were collected for all participants. In addition, participation in the study was dependent upon child attendance. Individual preference assessments were conducted in a small classroom adjacent to the participants' regular classroom. All group preference assessment procedures were conducted within the participants' regular classroom. During group preference assessment procedures, participants sat in their usual location within the classroom. The class was arranged such that students sat in rows of five students. Similar to Layer et al. (2008), all individual preference assessments were completed in one session, and all group preference assessments were completed in either one or two sessions, with no more than 24 h between each session.

Materials

Several materials were utilized during preference assessments, including edibles, the Plickers website, Plickers cards, an iPhone with the Plickers app, and a computer and projector.

Edible Stimuli Four small edible stimuli were utilized during all phases of the study: jelly beans, raisins, Sixlets, and baked cheese crackers. Specific edible stimuli were selected due to their distinct flavor profiles (e.g., sweet, fruity, malted, and salty), similar sizes, and utilization in the original study conducted by Layer et al. (2008). Additionally, participants had access to water during all sessions. Edible stimuli were selected for use as they could be delivered and consumed immediately, as well as for comparability with previous research (Layer et al., 2008). All edibles were approved by administrators and teachers prior to use.

Plickers Website The Plickers website (www.plickers.com) was utilized during all group preference assessment procedures. The website allows users to enter questions into a digital library for presentation to a group. The website also allows users to enter the names of group members and assign each member a unique Plickers card. By assigning each

member of the group a unique Plickers card, the website is then able to record individual responses and track changes in responding over time.

Plickers Cards Plickers cards were utilized during all group preference assessment procedures. Plickers cards are 4 in. × 4 in. cards that feature a QR (Quick Response) code, a type of two-dimensional barcode that allows for rapid readability (Fig. 1). A total of 63 Plickers cards, each with a unique QR code, are available for download from the Plickers website. In addition to featuring a unique QR code, each Plickers card features the letters A, B, C, and D in small print, with one letter on each edge of the four-sided square card. As the Plickers card is rotated such that a different letter is oriented at the top of the card, the QR code is read differently by the Plickers app—allowing each single Plickers card to be read as four different QR codes (i.e., code A, code B, code C, and code D), depending on how the Plickers card is rotated by the user. Given the unique QR code and small print of letters, use of Plickers allows for participants to respond without alerting anyone to their selection. The Plickers cards used in the current study were printed in black ink on white cardstock and laminated using matte lamination.

iPhone with Plickers App The Plickers app was downloaded onto an iPhone 5. The app is available at no cost. The Plickers app allows the user to select a question from the library to be displayed on a computer or digital overhead projector. Following projection of a question, group members rotate their Plickers card such that their preferred response (e.g., A) is at the top of the card. Next, the facilitator uses the Plickers app to scan the responses of group members. The app utilizes the user's smartphone camera to determine the orientation of the card of each group member. Scanning is accomplished simply by pointing the camera in the direction of the group, while the app records multiple group member responses simultaneously. The user of the app is not required to move toward individual group members to read a card, as Plickers cards can be read from in excess of 20 ft. As the user is scanning the classroom, transparent boxes with each group members name are displayed on the screen of the smartphone. As a group member's Plickers card is successfully scanned, the transparent box is filled with green—allowing the user to identify group members whose responses have yet to be scanned.

Computer and Projector A PC connected to an overhead projector was used during all group preference assessment sessions. The computer was used to log into the Plickers website. Once logged into the Plickers website, the overhead projector was used to display a website-based question, uploaded by the experimenter, which stated, "Which would you like," followed by A—raisin, B—cracker, C—jelly bean, and D—sixlet,

Fig. 1 Example of a Plickers card

each of which corresponded with a different edible. Letter-edible associations were consistent throughout the study.

Experimental Design

This study used a reversal design to evaluate differences in selection due to preference assessment procedures. Similar to Layer et al. (2008), two individual preference assessments were initially conducted to determine stability of selection under identical preference assessment procedures. Next, participants were exposed to the group procedure. Each condition was then repeated once. Although a total of 19 participants were included in the study, 3 participants missed both sessions of the group arrangement due to school absences and thus were excluded from analyses.

Dependent Variable and Data Analysis

The dependent variable assessed in this study was stimulus selection. The definition of stimulus selection varied by assessment procedure. In the individual assessment procedure, stimulus selection was defined as pointing to one stimulus in an array or verbally identifying a stimulus. In group assessment procedures, selection was defined as orienting the Plickers card to the preferred stimulus. Data analysis procedures were identical to those utilized by Layer et al. (2008). One level of analysis consisted of calculation of indifference scores for each participant for each session. Indifference scores are used to quantify the amount of selection variability within a session and were used to evaluate the speed with which preference was identified across preference assessment

modalities. Indifference scores were calculated by dividing the number of trials administered within a session by the number of stimulus restriction criteria met plus one (Layer et al., 2008). Restriction criteria consisted of: (1) four consecutive selections of the same stimulus, (2) a stimulus being selected four times more than all other stimuli, (3) two stimuli both being selected equally and four more times than all other stimuli, and (4) 20 consecutive trials had elapsed with no restriction criteria being met. The possible range of indifference scores was 3 (participant preference quickly identified) to 20 (participant was indifferent). For the individual and group preference assessments, indifference scores were calculated on an individual participant level.

Additionally, stimulus hierarchies were constructed following each session to determine if diverse preference assessment types resulted in similar or dissimilar preference hierarchies. Preference hierarchies were calculated following each session for each participant by dividing the number of selections of a stimulus by the number of trials for which that particular stimulus was available for selection and multiplying by 100%. Stimuli were then ordered from most preferred (i.e., highest percentage) to least preferred (i.e., lowest percentage). In accordance with Layer et al. (2008) procedures, if percentages were equal across stimuli, the stimulus selected last was assigned the higher rank. Results were compared across three criteria. First, hierarchies were compared to determine whether the first stimulus selected across assessment administrations was the same. Agreement was recorded when the same stimulus was identified as the same across preference assessment administrations. Second, hierarchies were compared to determine the similarity of the entire hierarchy across procedures. Agreement was scored when stimuli moved no more than one rank across assessment procedures. Finally, Spearman's rank order correlations and Pearson correlations were calculated to evaluate relationships between datasets.

Efficiency of the two procedures was assessed by dividing the total time elapsed per preference assessment by the total number of students assessed. Total time elapsed was calculated by measuring the number of minutes elapsed between the start and end of each preference assessment method. So, for the individual assessment, the timing began when the first student began his or her assessment and ended once the final student in the classroom had completed his or her assessment. For the group assessment, timing began when the Plickers cards were passed out and ended once all students' Plickers cards were removed. Calculation of time elapsed allowed for assessment of time required to perform tasks not directly associated with the preference assessment itself (e.g., resetting assessment between participants, escorting participants to assessment setting).

Procedure

Individual Preference Assessment The individual preference assessment procedure was identical to that used by Layer et al. (2008) and is similar to those described in other studies (DeLeon & Iwata, 1996). In this procedure, one participant was present during each assessment. Prior to the first individual preference assessment, participants sampled each stimulus while the researcher labeled each stimulus (e.g., “This is the cracker”). After sampling each stimulus, participants were introduced to the individual preference assessment. During each individual preference assessment, the researcher arranged the four stimuli in a half circle on a table in front of the child. The researcher then instructed the participant to select the stimulus they would like to consume (i.e., “Which would you like?”), following which the selected stimulus was delivered to the participant. The participant was then allowed 30 s to consume the stimulus. After 30 s, any remaining portion of the stimulus was discarded. Failure to identify a stimulus would have resulted in the researcher waiting 5 s before redelivering the prompt to select a stimulus, but this did not occur at any point in the study. Additional procedures were in place for instances in which a participant simultaneously pointed to two stimuli or made an unclear selection, but this also did not occur. Following selection and consumption of a stimulus, a new trial began and the participant was directed to select the stimulus they would like to consume.

All stimuli were presented for all trials until one of the aforementioned restriction criteria was met, at which point the stimulus meeting a criterion was removed from the array. Once a stimulus was removed from the array, the assessment procedure continued until restriction criteria were met for the remaining stimuli. If at any point criterion (4) was met, the preference assessment was terminated.

Group Preference Assessment Prior to conducting the first group preference assessment, participants were trained in use of the Plickers cards to indicate a preferred stimulus. Training consisted of didactic instruction by the researcher, modeling of orientation by the researcher, group practice of orientation to a Plickers selection (e.g., all participants instructed to orient the Plickers card so that A was at the top of the card), and feedback regarding incorrect orientation when necessary. Training in use of Plickers was completed in one 5-min session. Participants were then informed that they would use the Plickers cards to indicate a preferred stimulus and that the selections of the entire class would determine the stimulus delivered. Participants were told to refrain from engaging in verbal responses regarding their preferred stimulus and to only use the Plickers card to indicate a response. Following training, the question “Which would you like?”, followed by the four response options, was displayed through the projector via the Live View feature of the Plickers website. Once the

question was displayed, participants rotated their Plickers card so that their response (e.g., A) was at the top of the Plickers card. The researcher then used the Plickers app to scan participant responses from the front of the classroom. In order to ensure that all Plickers cards were scanned, the researcher would move from one side of the classroom while scanning the cards. Anecdotally, scanning the responses of all students in the classroom took approximately 30 s per assessment. The Plickers app calculated the percentage of students that selected each stimulus. These percentages were then used to assemble a spinner with different sized sectors corresponding to the percentages (National Council of Teachers of Mathematics, 2016), which was then displayed via the projector. Once sector sizes were determined, the spinner was spun and the stimulus identified by the spinner was delivered to all participants in the classroom, regardless of individual stimulus selection. Although not formally measured, the estimated time elapsed between stimulus selection and delivery of a stimulus was approximately 1 to 2 min. Anecdotally, participants occasionally expressed excitement or displeasure with the stimulus delivered.

Following delivery of the stimulus, a new trial was administered. Restriction of stimuli was tracked using the Plickers website, with stimuli being restricted on an individual participant level. That is, once an individual student met restriction criterion (1), (2), or (3), a researcher placed a sticker on the corresponding side of the participant’s Plickers card, signifying that the participant was no longer able to select that stimulus. If a participant oriented their Plickers card to a restricted side, a researcher reminded the participant that they were not able to select that stimulus and prompted the participant to select a different stimulus. Once a preference hierarchy was established or a participant met criterion (4), the participant’s Plickers card was removed from the system. The participant was allowed to continue to identify their preferred stimulus without restriction, but their selection was not recorded. The group procedure was terminated when all participants either had all stimuli restricted or met restriction criterion (4).

Reliability

Interobserver agreement (IOA) was assessed during 40% of individual preference assessments and 100% of group preference assessments. During individual preference assessments, IOA was calculated by dividing the number of agreements for participant stimulus selection divided by the number of agreements and disagreements and multiplying by 100. Mean IOA was 100% for the individual preference assessments. During both group preference assessments, two researchers converted data automatically collected by the Plickers app into cumulative frequencies for each participant. IOA was calculated by dividing the number of agreements regarding the cumulative frequency count for each participant by the number of

agreements plus disagreements. Mean IOA was 100% for the group preference assessment.

Procedural Integrity

Procedural integrity was assessed during each individual and group preference assessment procedure. Data were collected using study-derived integrity checklists unique to each condition. During each preference assessment, researchers recorded a procedural step as completed if the step was fully completed. Any step that was not fully completed was marked as incomplete. Procedural integrity was reported to be 100% for both individual and group preference assessments. IOA for procedural integrity was collected during 40% of individual preference assessments and 100% of group preference assessments, with IOA being 100%.

Results

This study sought to determine whether different preference assessment procedures resulted in differing levels of intra-session selection variability and whether different preference assessment procedures resulted in similar preference hierarchies. An example of trial by trial data, utilized to calculate indifference scores, is presented in Fig. 2 for two participants—one of whom identified the same stimulus as most preferred across assessments and one for whom highly indifferent responding was noted during individual preference assessments and highly discriminated responding was observed during group preference assessments. As an example of a calculation of an indifference score, participant 3's selections during the first individual assessment (Fig. 2) primarily varied between crackers and jelly beans, at which point the cracker and the jelly bean had been selected equally and four more times than all other stimuli participant 3—meeting restriction criterion (3). Participant 3 then selected Sixlets three consecutive times, resulting in restriction criterion (2) being met. The indifference score was calculated by dividing the number of total trials (i.e., 18) and dividing by the number of restriction criteria met (i.e., 2) plus 1, resulting in an indifference score of 6.0. Mean indifference scores across participants for each condition are presented in Fig. 3. Mean indifference scores were highest in the individual preference assessment. Standard deviations were also greatest in the individual preference assessment.

As the group preference assessment resulted in probabilistic delivery of participant selected stimuli, analyses were performed to determine the probability that a participant received the stimulus he or she selected. Individual participant probabilities for the group preference assessment procedure are presented in Table 1. A Pearson correlation was conducted to determine whether probabilistic delivery of stimuli in the

group arrangement was related to indifference scores. A non-significant correlation was found between indifference scores during the group procedure and the probability that a stimulus was delivered during that same procedure ($r = -0.05$, $p = 0.81$), suggesting that the probabilistic delivery of stimuli during the group procedure was not related to variable responding.

Analyses were conducted to determine consistency of first stimulus selected and overall hierarchies across all preference assessments (Table 2). When comparing individual preference assessments to other individual preference assessments, the same stimulus was selected first by a mean of 43% (range = 27–71%) of participants, with a similar hierarchy identified for a mean of 36% (range = 27–50%) of participants. Comparing the first group to the second group assessment, the same stimulus was selected first for 36% of participants and a similar hierarchy was identified for 36% of participants. Comparison across assessment types revealed identification of the same stimulus as most preferred for a mean of 40% (range = 18–54%) of participants, with a similar hierarchy established for a mean of 38% (range = 18–54%) of participants.

Similarity in preference hierarchies developed through each preference assessment was also evaluated through calculation of Spearman's rank correlation coefficient. Correlations for similar preference assessment procedures and different preference assessment procedures are presented in Tables 3 and 4, respectively. A post hoc power analysis indicated that the observed power ranged from 0.05 to 0.77 with an alpha of 0.05. As the number of days varied between comparison preference assessments, a Pearson correlation between days elapsed and rank correlations was conducted. A moderate, but statistically insignificant, negative correlation was found ($r = -0.46$, $p = 0.17$) for this comparison. Finally, relative efficiency of the two procedures was compared. Mean assessment time across the individual preference assessment conditions was 10.79 min per participant assessed. Mean assessment time across the group preference assessment conditions was 3.79 min per participant assessed. Although classrooms were referred for demonstrating high levels of disruptive behavior, anecdotal reports indicate high levels of compliance and limited disruptive behavior during both individual and group preference assessments.

Discussion

The primary purpose of this study was to conduct a systematic replication of Layer et al. (2008) by comparing the results of a group preference assessment procedure to individual preference assessments; however, unlike Layer et al. (2008), the group preference procedure was conducted with all students within the classroom simultaneously using an application

Fig. 2 Examples of cumulative selections and indifference scores across preference assessment procedures. Consistent participant responses across assessment types are depicted in the top panel. The bottom panel

depicts an example of highly indifferent responding during individual preference assessments and highly discriminated responding during group preference assessments

called Plickers. Similar to Layer et al. (2008), we found that the group preference assessment procedure resulted in lower indifference scores for a majority of participants. Specifically, ten participants exhibited a clear preference (i.e., indifference score < 20) for one of the four stimuli in the individual procedure. The group preference assessment procedure yielded lower indifference scores for all but two of these participants. Additionally, five more participants did not exhibit a clear preference (i.e., indifference score = 20) for stimuli during the individual preference assessment procedure but exhibited much lower indifference scores during the group procedure. These data imply that the group procedure encouraged more differentiated responding during the preference assessment process.

Similar to Layer et al. (2008), this study does not allow for determination of whether probabilistic outcomes or delay in delivery of reinforcement was responsible for lower indifference scores observed during group preference assessment procedures. Although use of Plickers provides an advantage over the procedure described by Layer and colleagues in that participant selection of stimuli is anonymous (i.e., participants are unable to identify selections of stimuli made by peers due to the unique nature of Plickers QR codes), it is still possible that peer influence may have influenced participant selection of stimuli. Although a vocal report of students' preference would have been more efficient it would not have retained the

Fig. 3 Mean indifference scores and individual participant indifference scores across assessment procedures

Table 1 Mean probability that a child received the same item selected

Participant	Probability
Participant 1	0.33
Participant 2	0.29
Participant 3	0.45
Participant 4	0.37
Participant 5	0.44
Participant 6	0.40
Participant 7	0.40
Participant 8	0.33
Participant 9	0.44
Participant 10	0.33
Participant 11	0.41
Participant 12	0.37
Participant 13	0.43
Participant 14	0.31
Participant 15	0.36
Participant 16	0.33

Table 2 Criterion agreements

Comparison	Ranking agreement	First ranked
Individual 1/individual 2	50%	71%
Individual 1/group 1	43%	43%
Individual 1/individual 3	27%	27%
Individual 1/group 2	46%	38%
Individual 2/group 1	54%	54%
Individual 2/individual 3	31%	31%
Individual 2/group 2	18%	18%
Group 1/individual 3	31%	46%
Group 1/group 2	36%	36%
Individual 3/group 2	38%	38%

privacy offered by the current methodology. Using Plickers, it was very difficult for students to determine which stimuli their peers were selecting. We chose to use this more private method to reduce the likelihood that students selecting less uniformly preferred stimuli (e.g., a raisin) were the subject of ridicule or peer pressure.

Unlike Layer et al. (2008), this current study found a moderate correlation between days elapsed and correlations between assessment procedures, indicating that differences in time between assessments may have contributed to variations in correlations. Whereas the longest span between assessment procedures in Layer and colleagues study was 69 days, the longest span in the present study was 111 days—which may account for differences in correlations. Although this finding differs from Layer and colleagues, it is somewhat unsurprising given research indicating the dynamic nature of preference (Ciccone, Graf, & Ahearn, 2007; Verriden & Roscoe, 2016).

The group preference assessment procedure holds utility within the context of class-wide group contingencies. As previously mentioned, group contingencies do not necessitate the delivery of a single stimulus as reinforcement to all members of the group. However, for practical reasons, they are often implemented in this way. Although the evidence supporting group contingencies support them as effective behavior change strategies in the absence of preference assessments, it is possible that some students do not contribute toward satisfying the contingency. A lack of participation may be driven by disinterest in the stimulus selected for reinforcement. Thus, instead of relying on teacher judgment or majority rule voting to select a stimulus,

the probabilistic delivery of stimuli as reinforcers can be easily transferred to the context of intervention. That is, teachers can generate a physical or digital spinner that reflects the results of the group preference assessment. The spinner could then be used to select a single stimulus for reinforcement each time the contingency is satisfied. In addition, weekly group preference assessments could be conducted to ensure the spinner accurately reflects the preferences of the group. Furthermore, teachers might use Plickers to assess group preference more frequently (e.g., daily) without generating a complete hierarchy. This application might be the most practical within classroom settings where group preference may shift more rapidly and complete hierarchies may not be as useful. Finally, it might be necessary to conduct additional assessment of sub-group preference if one or more students within the whole group do not modify their behavior to conform to the contingency. Although the group preference assessment procedure utilized in the current study was not overly time consuming, practitioners should carefully consider demands that frequent use of these procedures may place on end users in terms of materials and time, with future research being needed to evaluate the relative costs and benefits of such procedures.

Although less stability in preference hierarchies was noted in this study within and across preference assessment types in comparison to Layer et al. (2008), similar levels of identification of the most preferred stimulus across procedures were noted. Whereas Layer and colleagues documented statistically significant correlations between different preference assessment procedures, this study only found statistically significant correlations between repeated administrations of the individual preference assessment procedure. These correlations were moderate across a relatively short period of time (i.e., 5 days) and small across longer spans (i.e., 90 days), suggesting that student preference was variable. This is not inconsistent with previous research demonstrating instability in individual preference both within and between assessment methods (Call, Trosclair-Lasserre, Findley, Reavis, & Shillingsburg, 2012; Mason, McGee, Farmer-Dougan, & Risley, 1989; Zhou, Iwata, Goff, & Shore, 2001). These data highlight the importance of frequent assessment of stimulus preference across individual, and group methodologies and researchers should continue to investigate solutions to address these fluctuations across settings.

Conducting the group preference assessment using Plickers allowed all students in the classroom to respond rapidly to

Table 3 Correlations and days expired between assessments, same type

	Individual 1/individual 2	Individual 1/individual 3	Individual 2/individual 3	Group 1/group 2
Correlations between assessments	0.49*	0.19	0.28*	0.26
Mean number (range) of days between assessments	5 (0–20)	94 (90–98)	90 (90–91)	25 (25–25)

*Statistically significant at $p < .05$

Table 4 Correlations and days expired between assessments, different types

	Correlations between assessments	Mean number (range) between assessments
Individual 1/group 1	0.20	22 (0–28)
Individual 1/group 2	0.01	102 (98–106)
Individual 2/group 1	0.21	20 (19–21)
Individual 2/group 2	–0.16	97 (84–99)
Individual 3/group 1	0.32*	71 (70–71)
Individual 3/group 2	0.08	5 (0–8)

*Statistically significant at $p < .05$

assessment trials and substantially decreased the total time taken to conduct the group preference assessment compared to the individual assessment method. The use of Plickers as a means of collecting participant selection data may further reduce logistical challenges associated with the use of alternative formats for selection of stimuli (e.g., cards, poker chips) within a group preference assessment procedure. Utilization of Plickers in this study allowed the researcher to stand at the front of the room and quickly collect all participant responses via the Plickers app without having to move about the classroom or ask students individually regarding preference. In addition to allowing participants to respond anonymously, use of Plickers allowed researchers to automatically record selection data at the individual student level. In an applied setting, student-level preference data could be beneficial in determining why an individual student is failing to respond to a group contingency (e.g., stimuli that are not highly preferred by an individual student are repeatedly delivered as rewards), allowing reinforcement procedures to be tailored to address non-responders within a group. Whereas other studies have utilized surveys of possible reinforcers, Plickers also improves on this method of group preference assessment by eliminating the need for hand calculation of preferred stimuli. Furthermore, because Plickers are freely available, require little training to ensure correct usage of Plickers by students, and are versatile in their application to behavior analysis, future research should focus on how they may be incorporated into other practices in the field.

Limitations

The results of this study must be viewed in light of several limitations. First, all participants experienced preference assessments in the same order (i.e., AABAB). As such, it is possible that the lower variance in stimulus selection within the group preference assessment procedures was due to initial experience with stimuli during the individual preference assessment; however, it must be noted that participants returned to higher levels of indifference during the third individual preference

assessment as compared to the preceding group preference assessments (e.g., Fig. 3). Second, no reinforcer assessment was conducted as part of this study. As such, it is unknown whether the stimuli identified as preferred in either the individual or group preference assessments functioned as reinforcers capable of modifying participant behavior within a group contingency. Future researchers should investigate the utility of such procedures in identifying effective reinforcers for appropriate classroom behavior. Additionally, researchers need to evaluate the utility of the group preference assessment procedure within an intervention context. Specifically, we must consider direct comparisons between group contingencies that are informed by preference assessment and those that are not before recommending widespread use and utility of these procedures.

Third, all stimuli available for selection in the current study were edibles. Edible stimuli were selected for use in the study due to procedural similarity with Laver et al. (2008) and due to the fact that similar rewards were already utilized in the classroom. As school settings may utilize activities or tokens as rewards within a group contingency (Filcheck, McNeil, Greco, & Bernard, 2004; Kehle, Bray, Theodore, Jenson, & Clark, 2000), future researchers should evaluate the use of non-edible stimuli, as these stimuli may be associated with longer delays between selection and contact with a reinforcing stimulus. Fourth, participants in the current study were limited to one general education middle school classroom. As such, it is unknown whether dissimilar populations (e.g., younger students, students of different functional skill levels) would be able to utilize Plickers as easily as the students in the present study. Therefore, it is essential that these procedures be evaluated with other populations. Fifth, although students were instructed to refrain from sharing their preferences during the group assessment procedure, we did not collect systematic data to evaluate that aspect of the assessment. Anecdotally, we did not observe students sharing preferences for stimuli and the Plickers cards made it difficult to do so; however, it is entirely possible that students shared this information with each other leading to increases in discriminated responding in the group format.

Sixth, we did not formally assess the reliability of the Plickers app's ability to scan the QR codes accurately. Although it is highly unlikely that participants' responses were incorrectly identified, it is possible that the app erred in this regard. Finally, there were some procedural differences between the two types of preference assessments that may have impacted our findings. Specifically, the individual preference assessment had participants selecting and consuming an edible item to indicate preference. On the other hand, the group preference assessment procedure had participants indicate their preference using Plickers cards and receiving the edible stimulus after a short delay. Equating the two (i.e., having participants respond with Plickers cards regardless of methodology) would have decreased the procedural differences, and it is not known how these differences impacted the present findings.

Conclusion

This study adds to the literature supporting the use of group preference assessments procedures (Layer et al., 2008) with findings indicating that a group preference assessment procedure is more efficient and decreases student indifference for sample stimuli. Additionally, this study provides a model use of Plickers as a means of selection of preferred stimuli, allowing for efficient collection of selection data of each participant within a group. Although the findings of the current study indicate that use of Plickers may be promising, researchers and practitioners should carefully consider the population with whom they implement such procedures as participants were limited to general education middle school students. Finally, the current study provides important data regarding procedures that may be utilized to evaluate preference within a group setting; however, the ability of stimuli included to modify behavior ultimately depends upon the reinforcing value of those stimuli. As such, careful consideration must be made when determining possible stimuli to ensure that stimuli that are reinforcing for all individuals of a group are made available for selection.

Compliance with Ethical Standards

All procedures performed in the current study were in accordance with the ethical standards of the institution and the national research committee and with the 1964 Helsinki declaration and its later amendments.

Informed consent and assent was obtained for all individual participants included in the study.

Conflict of Interest The authors declare that they have no conflict of interest.

References

- Barrish, H. H., Saunders, M., & Wolf, M. M. (1969). Good behavior game: effects of individual contingencies for group consequences on disruptive behavior in a classroom. *Journal of Applied Behavior Analysis*, 2, 119–124.
- Bradshaw, C. P., Reinke, W. M., Brown, L. D., Beavans, K. B., & Leaf, P. J. (2008). Implementation of School-Wide Positive Behavioral Interventions and Supports (PBIS) in elementary schools: observations from a randomized trial. *Education and Treatment of Children*, 31, 1–26.
- Call, N. A., Troscclair-Lasserre, N. M., Findley, A. J., Reavis, A. R., & Shillingsburg, M. A. (2012). Correspondence between single versus daily preference assessment outcomes and reinforcer efficacy under progressive-ratio schedules. *Journal of Applied Behavior Analysis*, 45, 763–777.
- Ciccone, F. J., Graff, R. B., & Ahearn, W. H. (2007). Long-term stability of edible preferences in individuals with developmental disabilities. *Behavioral Interventions*, 22, 223–228.
- Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis* (2nd ed.). Upper Saddle River: Pearson Prentice.
- Dart, E. H., Radley, K. C., Battaglia, A. A., Dadakhodjaeva, K., Bates, K. E., & Wright, S. J. (2016). The classroom password: a class-wide intervention to increase academic engagement. *Psychology in the Schools*, 53, 416–431.
- DeLeon, I. G., & Iwata, B. A. (1996). Evaluation of a multiple-stimulus presentation format for assessing reinforcer preferences. *Journal of Applied Behavior Analysis*, 29, 519–532.
- Filcheck, H. A., McNeil, C. B., Greco, L. A., & Bernard, R. S. (2004). Using a whole-class token economy and coaching of teacher skills in a preschool classroom to manage disruptive behavior. *Psychology in the Schools*, 41, 351–361.
- Gresham, F. M., & Gresham, G. N. (1982). Interdependent, dependent, and independent group contingencies for controlling disruptive behavior. *The Journal of Special Education*, 16, 101–110.
- Hawkins, R. O., Haydon, T., Denune, H., Larkin, W., & Fite, N. (2015). Improving the transition behavior of high school students with emotional behavioral disorders using a randomized interdependent group contingency. *School Psychology Review*, 44, 208–223.
- Heal, N. A., & Hanley, G. P. (2007). Evaluating preschool children's preference for motivational systems during instruction. *Journal of Applied Behavior Analysis*, 40, 249–261.
- Hirsch, S. E., Healy, S., Judge, J. P., & Lloyd, J. W. (2016). Effects of an interdependent group contingency on engagement in physical education. *Journal of Applied Behavior Analysis*, 49, 975–979.
- Kehle, T. J., Bray, M. A., Theodore, L. A., Jenson, W. R., & Clark, E. (2000). A multi-component intervention designed to reduce disruptive classroom behavior. *Psychology in the Schools*, 37, 475–481.
- Kodak, T., Fisher, W. W., Paden, A., & Dickes, N. (2013). Evaluation of the utility of a discrete-trial functional analysis in early intervention classrooms. *Journal of Applied Behavior Analysis*, 46, 301–306.
- Krentz, H., Miltenberger, R., & Valbuena, D. (2016). Using token reinforcement to increase walking for adults with intellectual disabilities. *Journal of Applied Behavior Analysis*, 49, 745–750.
- Lannie, A. L., & McCurdy, B. L. (2007). Preventing disruptive behavior in the urban classroom: effects of the Good Behavior Game on student and teacher behavior. *Education and Treatment of Children*, 30, 85–98.
- Layer, S. A., Hanley, G. P., Heal, N. A., & Tiger, J. H. (2008). Determining individual preschoolers' preferences in a group arrangement. *Journal of Applied Behavior Analysis*, 41, 25–37.
- Maggin, D. M., Johnson, A. H., Chafouleas, S. M., Ruberto, L. M., & Berggren, M. (2012). A systematic evidence review of school-based group contingency interventions for students with challenging behavior. *Journal of School Psychology*, 50(5), 625–654.
- Mason, S. A., McGee, G. G., Farmer-Dougan, V., & Risley, T. R. (1989). A practical strategy for ongoing reinforcer assessment. *Journal of Applied Behavior Analysis*, 22(2), 171–179.
- National Council of Teachers of Mathematics (2016). Adjustable spinner. Retrieved from: <https://www.nctm.org/Classroom-Resources/Interactives/Adjustable-Spinner/>.
- Piazza, C. C., Roane, H. S., & Karsten, A. (2011). Identifying and enhancing the effectiveness of positive reinforcement. In W. W. Fisher, C. C. Piazza, & H. S. Roane (Eds.), *Handbook of applied behavior analysis* (pp. 151–164). New York, NY: Guilford Press.
- Roane, H. S., Vollmer, T. R., Ringdahl, J. E., & Marcus, B. A. (1998). Evaluation of a brief stimulus preference assessment. *Journal of Applied Behavior Analysis*, 31, 605–620.
- Verriden, A. L., & Roscoe, E. M. (2016). A comparison of preference-assessment methods. *Journal of Applied Behavior Analysis*, 49, 265–285.
- Wehman, P. (1976). Selection of play materials for the severely handicapped: a continuing dilemma. *Education and Training of the Mentally Retarded*, 11, 46–50.
- Zhou, L., Iwata, B. A., Goff, G. A., & Shore, B. A. (2001). Longitudinal analysis of leisure-item preferences. *Journal of Applied Behavior Analysis*, 34, 179–184.