Evaluating the Impact of Token Economy Methods on Student On-task Behaviour within an Inclusive Canadian Classroom

Robert L. Williamson, Chelsea McFadzen
Simon Fraser University, Canada

Abstract

A token economy is a common classroom positive behaviour support method whereby 'tokens' are delivered to students contingent on exhibiting specific behaviours. Students later exchange earned tokens for items of interest. This project developed a prototype, iPad-based tool that enabled teachers to deliver and track tokens virtually. The virtual token economy system was then compared to implementation using a typical, physically delivered token economy method. Both methods were evaluated concerning their impact with regard to grade four-five student's on-task behavior within one inclusive Canadian classroom using a multielement design. Individual impacts and group effects were analyzed using an analysis of variance with planned contrasts as well as visually utilizing single case methods to assess efficacy regarding each implementation method. Results indicated that only one significant difference for one individual subject was found between baseline (no token economy) and both token economy systems. No other significant differences were found between individual or group on-task behaviours nor between the baseline, physical and virtual methodologies overall. Implications regarding evidence that TEs represent evidence-based practice and suggestions for future research are discussed.

1. Introduction

A token economy (TE) is a secondary reinforcement system of positive behaviour support whereby tokens (i.e., conditioned reinforcers) are delivered to students for exhibiting specific behaviours [1, 2, 3, 4, 5]. These tokens represent a medium of exchange to be used by recipients to purchase desired goods or privileges from a menu of items [6, 2]. TE systems have been used in a variety of settings and over many decades within an academic environment [1, 2, 7]. Over a decade ago, TEs were identified by Simonsen and colleagues (2008) as meeting criteria for evidence-based practice and by the American Psychological Association's Task Force on Promotion and Dissemination of Psychological Procedures (1993) as a well-established psychological procedure.

With a long history of use within academic and other settings, the TE has enjoyed a reputation as an evidence-based classroom behaviour management tool and has widely been considered effective in decreasing non-desired behaviours and increasing pro-academic behaviours in students [8, 9, 4, 7]. Studies have also shown that token economy systems have been used to increase on task behaviours and decrease non-desired behaviours [10, 11].

Some however, have questioned the assertion that the TE should be considered an evidence-based practice. Maggin, Chafouleas, Goddard and Johnson [12] conducted a systematic evaluation of research involving TEs as classroom management tools for students with challenging behaviours and found that the "…extant research on token economies (did) not provide sufficient evidence to be deemed best practice based on the WWC (What Works Clearinghouse) criteria" [12]. Authors suggested that this finding was largely due to inadequate research designs in their uncovered literature. Specifically, the authors cited a lack of information within studies regarding treatment fidelity and social validity as among the methodological problems found in the literature available at the time of their systematic evaluation [12]. This finding presents a profound concern, as without sufficient description of the exact procedures used within literature that finds positive or negative results, the aggregate effectiveness of any given implementation method of a TE cannot be appropriately assessed.

In a more recent meta-analysis by Soares, Harison, Vannest and McClelland [7] conducted five years after Maggin et al., the use of a token economy in a classroom setting was found to "…suggest that a TE is an effective intervention, specifically for use in the classroom setting" [7]. Unlike Maggin et al.'s finding

that only 30% of studies were rated as achieving a medium to strong quality design based on WWC standards, Soares et al. found that 64% met that same WWC criteria as either medium or strong and suggested that the quality of research regarding the effectiveness of TEs is improving. Still, Soares et al. commented that "…only a third of the studies reported treatment fidelity." [7]. Clearly, differences remain in evaluating this intervention.

Ivy et al. [2] conducted a systematic literature review regarding the quality of procedural descriptions within TE research and found that "…of the 96 articles reviewed, only 18 (19%) included procedural descriptions of each component to a degree sufficient to guide replication" [2]. This finding would seem to support the previous assertions of Maggin et al. [12] in noting a lack of adequate implementation information regarding the specific TE methodologies applied within published studies. Differing findings leads to questions as to exactly how teachers are implementing TE's evaluated in previous studies and if any specific attributes of implementation are more or less impactful upon specific student behaviours.

There are six essential components of a TE system [13, 2, 14]. These six components include (1) the target behavior that is the focus of the intervention, (2) the tokens themselves, which must have been conditioned to function as reinforcers, (3) the backup reinforcers that may be purchased with a token, (4) the method by which tokens are earned, (5) the method by which tokens may be exchanged for backup reinforcers, and (6) the cost of the backup reinforcers [2].

Traditionally, TE's have been implemented within classrooms using physical tokens that are delivered to students. Teachers carry tokens (often fake money, poker chips or similar type of token) on their person as they teach. When a student displays the desired behaviour, the teacher delivers a token to the student. This physical method requires physical proximity to the student receiving the token and delivering the reward in person from teacher to student. This requirement may interrupt teacher instructional leadership.

The goal of this research was to examine any relative impacts concerning student on task behaviour between three separate conditions. Condition one consisted of baseline (no TE implementation) while the remaining two consisted of 1) a prototype virtual iPad-based TE methodology known as 'CARS' (Class-wide Augmented Reward System) and 2) a traditional, physically implemented TE system. Within both method (iPad and physical), specific implementation methods were defined and followed. Both utilized a variable ratio reward system of token delivery and the observation of interest was student on-task behaviour.

2. Method

2.1. Setting

Participants were recruited from one typical inclusive elementary school classroom located in a non-urban area within the lower mainland of British Columbia (BC), Canada. A grade 4-5 combined inclusive classroom was chosen by the district based on teacher and school interest in the study. The classroom was located on the second floor of a single school building, at the end of a long hallway. Inside the classroom, tables were generally oriented in the first two-thirds of room space while the front third contained a crescent table (for group work) on the left, a carpet just under a smart board in the center, and a teacher desk in the front right of the room. The back of the room contained cabinets over the length of that wall. The wall adjacent to the back wall and opposite the door was fully windowed. Supplies were placed on lower cabinets along the windowed wall and iPads were stored and charged in the corner on the lower cabinets between the back and windowed walls. No TE system had been in use within the selected classroom prior to initiation of this study.

2.2. Participants

This mixed grade 4-5 class consisted of 23 total students within grades four (n=13) and five (n=10). The class was taught by one Caucasian female, BC licensed professional teacher with four years total teaching experience. The classroom employed one full time, one-on-one education assistant. Two students were designated by the district as having chronic health conditions and seven students were designated as having behavioural difficulties using BC Ministry of Education disability determination guidelines. All students gave assent to participate in the study and parent/guardian permissions per ethics review board protocols were obtained. The teacher likewise gave consent to participate in the research.

Three students were selected by the teacher for individual observation. Bob (pseudonym) was in grade four. Bob was a Caucasian male and was designated by school personnel as having a behaviour disability and a chronic health condition as defined by the BC Ministry of Education. Hellen (pseudonym) was a grade five Caucasian female and was designated by school personnel as having a behaviour disability as defined by the BC Ministry of Education. Mark (pseudonym) was a grade four Caucasian male and

was also designated by school personnel as having a behaviour disability as defined by the BC Ministry of Education. All students were native English speakers and participated in the typical BC standard curriculum for 100% of the school day. For each of the three students of specific interest, the behaviour of concern was identified by the teacher as time on task and thus this was the behaviour of observation in the present study.

2.3. Intervention Agent and Training

The teacher represented the intervention agent for this work. The role of the teacher as intervention agent consisted of 1) learning how to implement a TE, 2) teaching students how to engage in a TE, 3) initiating both the physical and virtual TE methods on preset days and phases of data collection and 4) adhering to the academic schedule during implementation and providing for reward redemption on days of implementation. The teacher was trained by the first and second authors regarding the six principles of a TE via individual one-on-one training regarding specific implementation factors in the classroom setting. Efficacy of the teacher training was assessed by observing the teacher's instruction regarding the TE to her students (as implemented in the classroom) by the first and second authors. The teacher covered all six aspects in a functional way with the students during the introduction of the TE with the students. The teacher was then observed during trial runs of both the physical and virtual implementation methods in her classroom. Implementation fidelity was observed via an implementation fidelity requirements list (see appendix A). Trial runs showed that the teacher understood and implemented both TE methods as required by the six components of a TE noted and adhering 100% to the implementation fidelity checklist as noted independently by the authors. No additional qualification nor training was deemed necessary nor provided prior to research data collection implementation. Follow up training after initiation of the TE research protocols was likewise not required as implementation protocols during implementation phases did not deviate from the implementation requirements.

2.4. Materials

All students were provided with one 9.5-inch iPad containing the student version of the CARS app each. The teacher was provided with a 12.9-inch iPad pro containing the teacher version of the CARS app.

The CARS system consisted of two interconnected iPad apps. The teacher 'signed up' each student in an online class portal. Students were then able to securely log in to their individual student app on their individual student iPad. The teacher likewise securely signed in on the teacher app from the teacher iPad. The student app allowed students to view tokens already obtained (a bank), prizes available and token price of each, as well as showed when a token was awarded via a 'pop up', push-type individual text notification message. The pop-up notification worked similar to all text messages on the iPad and thus the student app did not need to be activated in order for the pop-up notification to appear. The student could also be working on a different app on their iPad and the pop-up notification of an awarded token would still appear.

The CARS prototype virtual application was designed to mitigate possible struggles related to physically delivering tokens and gathering data by utilizing a specially designed, prototype iPad software tool. Specifically, the prototype software tool was designed to mitigate two difficulties teachers may face when implementing a token economy: 1) The iPad-based tool eliminated the need to physically deliver a token to a student. Instead, the teacher delivered tokens virtually by tapping on the picture of a student on the teacher iPad. Alternately, the teacher could award the whole class tokens via a 'whole class' button on the teacher app. It was hypothesized that this would improve temporal contingency relating behaviour to receipt of a token, save instructional time and minimize modest disruption when a teacher using a traditional physical TE might have been required to disengage from an instructional activity to deliver a token physically. Virtual token delivery also became a private rather than a public event when the student's iPad software recorded delivery and delivered the individual pop-up text message to the student's iPad confirming token receipt. 2) The iPad-based tool automatically recorded token delivery time and amount as data that was then available to the teacher on the system's web-based portal. The iPad also recorded tokens exchanged by the student, what they were exchanged for, and when the exchanges occurred within the same portal. Although outside the focus of this present study, such data could then be analyzed at a later time by the teacher to validly adjust token exchange intervals or reward choices for individual students at the time and discretion of the teacher.

3. Methodology

This research utilized a multi-element design. A single case alternating treatment (ABCBC) design was used to visually investigate the efficacy of two different versions of the token economy classroom management strategy upon baseline student on task behaviours. Baseline data (A) was taken in absence of any TE system of behaviour support in place. Then the

token economy method was implemented under two conditions: B) traditional (physical) token delivery and C) the prototype iPad-based virtual token delivery. On task behaviour was defined based on a related definition from Lee, Sugai and Horner [15] as a student that exhibits engagement of his/her senses and focus on the activity of instruction indicated by the teacher at the momentary time sampled. Student actions such as pausing, sleeping, prolonged gaze in a non-relevant direction, engaging or remaining disengaged from communication depending on the instructional activity and/or engaging in any non-relevant activity was an indication that the student was not reasonably attending to the instructional task.

3.1. Procedures

Data collection and research protocol implementation was scheduled and took place during morning academic activities. Each morning, students first engaged in whole group (class-wide) instruction led by the teacher. During whole group instruction, the teacher frequently sat on a stool located on a small carpet and within easy access to a classroom smart board. During this time, students were able to choose to sit in chairs or on a carpet during the whole group instructional activities. Whole group attention to task data collection was conducted during whole group instruction activities and took place at this same pre-determined and routine timeframe of the classroom schedule. No individual student data was collected during the whole group activities.

Following whole group instruction, students attended recess for approximately 15 minutes. Upon returning to the classroom, students engaged in stations-based instruction. The stations were located within the classroom (and one station sometimes located just outside the open classroom door at a hallway table). During stations work, the teacher led instruction in reading development activities from one of the stations (typically 3 or 4 total stations in operation during the stations activities) by sitting behind the crescent shaped table with her orientation out toward the class and students from the forward left-hand corner of the room. It was possible for the teacher to see all students during stations instruction (except any student engaged in activities using the hallway table just outside the classroom door). Other stations not led by the teacher were structured as independent learning activities for students at those stations. The classroom educational assistant generally monitored student activities at the stations as well as individual students during station activities in the classroom. Students rotated as cued by the teacher from station to station throughout the hour. During the stations-based instruction, data was collected regarding the three individual students of interest and not regarding the whole group. Both hours of instruction focused on language arts and reading related activities. This identical schedule of activities was followed each day that data was collected.

3.2. Data Collection

A momentary time sampling methodology [16] was implemented by designating multiple 15-minute periods over the course of each two hours +/- of data collection per day. During the 15-minute intervals within the first hour of whole class instruction, observations concerning the on or off task behaviour of the entire group of students was obtained using a timed camera snapshot of the students at the end of each minute of the 15-minute interval. Two cameras (for accuracy of angle and inter-rater review purposes) were placed high up in the front right and left corner of the rectangular room in a way as to capture the activities of all students within the room when the picture was snapped. Pictures were snapped automatically and without human interaction with the cameras via a commercially purchased app designed for that purpose. The snapping of pictures did not make a sound, nor did it make any visually observable action so as not to divert any student attention from the lesson/task being taught. Additionally, an independent data recorder (one or both authors) observed the group directly and noted the activities to which the students were to be engaged during that time. Following the 15-minute period, pictures were analyzed to count how many students were focused on the instruction or engaged in a directed activity for that sample captured in the photo based on the previously described definition of on-task behaviour. Data points for group on-task behaviour were then calculated by dividing the number of students on task for each sample picture (15 pictures in 15 minutes) by the total number of students within each picture frame. One data point was then calculated as percent on task over the entire 15-minute period by averaging the individual picture data points taken in the 15-minute period for the group of students. The unit of analysis was the average on-task percent over a one 15-minute period.

During physical TE implementation and during whole group instruction, the teacher utilized variable ratio (slot machine) reward schedules to deliver tokens by physically handing ‘toy dollar bills’ as tokens to students paying attention. A variable ratio schedule of token delivery is generally accepted as effective regarding the reinforcement of on task behaviours [17, 18]. At times, the teacher would hand a bill to all students by walking around the room as students engaged in a directed activity related to the whole group instruction. The use of variable ratio reward

(token) distribution was requested by the teacher in order that any interruptions to the flow and pace of the intended instruction would be minimized. Students would be asked to keep the bills in an envelope until access to their personal items (such as backpacks or notebooks) were accessible.

During virtual implementation, the teacher would award tokens through tapping a picture of an individual student or tapping the group button on the teacher iPad during whole group instruction while maintaining a variable ratio reward schedule. All students were located within visual proximity to their individually assigned iPads during virtual implementation, as the teacher announced the start of the virtual TE implementation prior to whole group teaching by asking all students to retrieve their individually assigned iPads and sign in prior to lesson initiation.

During the second hour of instruction (stations/small group and independent activities), each of three pre-selected students were observed during a 15-minute period using a momentary time sampling methodology using the same definition of on-task behaviour. Data was collected by the first and/or second authors by observing each of the three students at the end of each minute of each 15-minute period and noting if the student was on or off task relative to the educational activity assigned. This was then converted to a percentage on task by dividing the number of points of on task observations by the total number of observations in the period (15) and multiplying by 100. A single percent on task data point was recorded that represented one student over the entire 15-minute period of observation per student. The unit of measure for individual student on-task behaviour was one data point representing the average on-task percentage of the student over one 15-minute period.

During stations work, physical implementation of the TE method was conducted by the teacher through assigning individualized tasks to students at the station in which the teacher was leading instruction and then physically 'roaming' the room handing out bills to those students on task. The teacher also awarded bills occasionally to the students assigned to her own station. During virtual implementation, the teacher remained at the station in which she was directing instruction and gave tokens electronically to students on task by visually (and not physically) observing students in the room. A variable ratio token delivery schedule was used for both the physical and the iPad-based methodologies. Additionally, at least one token was delivered to one student (as a minimum requirement) over each group and individual 15-minute observation time period.

3.3. Inter-rater reliability

Inter-rater reliability (IRR) was conducted on 13 of 31 (32.2%) individual data collection sessions (each containing 15 separate data points to compare) by collecting data on the three individual students of interest by both the first and second authors simultaneously. After independent collection, data points were compared and percent agreement over each data point within each 15-minute period for each individual student (of the three targeted students) and a percent agreement was calculated. IRR achieved an average of 89.96% agreement (range: 82.2%-97.7%) for observations of the three individual students in total. To conduct IRR on the whole group attention to task data, the pictures were analyzed independently by the first and second authors. Group IRR was conducted on 5 of 14 sets of 15 pictures each or 35.7% of total observed data points and achieved 93.72% agreement.

3.4. Implementation Fidelity

Prior to implementation of the physical and virtual token economy systems, the participating teacher was trained in how to implement both forms of the TE systems. Practice with each form (physical and virtual) was conducted with feedback given to the teacher by the authors. Following teacher training and practice, the teacher relayed the method to the students in the class and was observed as accurate in describing the process to the students by study authors. The teacher explained that 1) 'paying attention' to lessons and activities was the desired behaviour. Teacher role played attention vs. inattention with specific reference to where one's eyes were looking vis-à-vis lesson involvement. Regarding tokens, the teacher explained that when she noticed students paying attention, she would award a token (individually) and if she noticed the group paying attention, she would award all of them a token. This was exemplified by asking students to perform an activity as the teacher went to each student noting if and how the student was paying attention providing specific feedback to each as she handed the student a token. The students were surveyed by the teacher to obtain reasonable 'prizes' that students could redeem tokens to obtain. The students brainstormed prizes and together with the teacher, listed those that would be available and at what price. This menu was posted in back of the room on a corner cabinet door that was used for prize redemption and within individual CARS student iPad apps. Tokens were to be redeemed at recess periods, lunch or after school. The recess period occurred directly in between the two hours of data collection/method implementation. Students

understood that these were the only times in which prizes could be redeemed.

These initial preparations adhered to the six vital components of TE methodologies noted by Ivy et al. [2] in the following ways. 1) Students were trained by the teacher and under the observation of study authors as to exactly what behaviour constituted reward of a token. 2) Students knew the value of each token by participating in the development of the different rewards and the costs related to each. According to the teacher, all students cognitively understood relative value and participated in choosing items/activities they valued to be placed on the menu of rewards. 3) Students understood that both 'real' and 'virtual' tokens could be combined to purchase items from the menu. All students were capable of independent mathematics required to add the physical and virtual tokens together. 4) Students understood that tokens were being given only during the two hours of observation in the mornings in which the data recorders (first and/or second author) were in the room observing and taking data. The tokens were not given on a specific schedule but instead were given according to a variable ratio method by the teacher as time and teaching methodology permitted her to note the attentive behaviour of individual or groups of students. Tokens were given no less than once during each 15-minute period to at least one student. 5) Students understood that recess, lunch and after school were designated as times that tokens could be redeemed based on teacher availability. Rewards were available during at least one of the times each day of study observation/implementation. 6) The menu of rewards contained the costs for each. Last, the implementation fidelity checklist was used to ensure the teacher adhered to these mandates of implementation each day of data collection achieving 100% adherence.

3.5. Data analysis

Regarding whole group attention to task, all data for all students' percent on task was calculated between the three conditions. A one-way, independent samples analysis of variance (ANOVA) was used to analyze any difference between or within percent time on-task among phases. Similarly, a one-way, independent samples ANOVA was conducted for each of the three students that were the focus of individual behaviour support to examine any difference between each student's on-task performance among the three conditions. Last, each condition and set of student data was examined using single case visual analysis techniques.

4. Results

Data was collected across eight total days between April 30, 2018 and June 4, 2018. Some differences exist concerning total number of 15-minute sessions (data points) between whole group data and each of the three individual student observations. This is due to absences from the class for any given student thus impacting total available time to observe and collect data for that student. The following results step through each planned comparison of means and visual inspection process.

Regarding any differences in whole group student on-task behaviours between the three conditions (no intervention, physical method, virtual method), data included all picture-based analysis of whole group activities. No significant difference was detected [$F(2,15)=2.211$, $p< .05$] between any of the phases. Planned contrasts showed that the implementation of either of the two methods (physical and virtual) did not significantly differ from baseline (no TE) [$t(15)= -1.413$, $p<.05$ (one tailed)] and that virtual implementation did not significantly differ from the physical implementation [$t(15)=-1.557$, $p<.05$ (two tailed)] regarding whole group on-task behaviour. Visual single case analysis of whole group data similarly did not reveal notable trends between phases (see Figure 1).

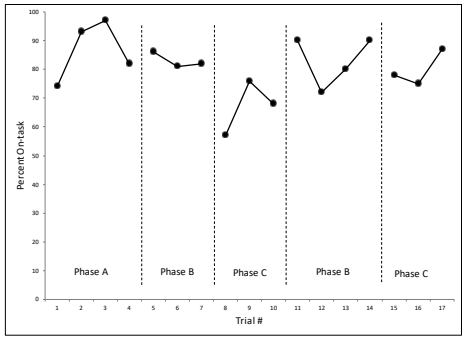

Figure 1. All Students

Regarding Bob (pseudonym), a one-way analysis of variance was calculated to test if the mean instances of on-task observations differed significantly between any of the three phases at the $p<.05$ level. Results indicated that there was a significant and large effect of the TE (not any specific version) on the target on-task behaviour of Bob [$F(2,20)=4.375$, $p<.05$, $\omega=.48$]. Using a Tukey HSD post hoc analysis, the differences in means between baseline and the virtual implementation was significant ($p<.05$). Further analysis using planned contrasts revealed that the significant effect was shown between baseline and the

TE implementation (both physical and virtual combined) [$t(20)=2.954$, $p<.01$ (two tailed) but did not indicate a significant difference between virtual and physical implementation methods [$t(20)=.029$, $p<.05$]. Visual single case plot analysis confirmed a positive difference between baseline and TE implementation phases but did not exhibit trends between the two implementation phases themselves (see Figure 2).

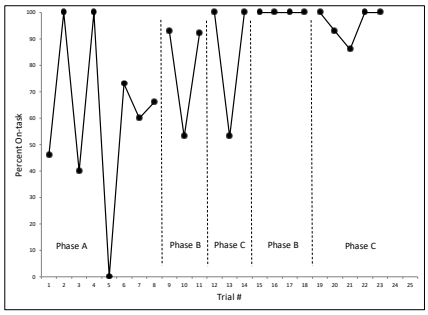

Figure 2. Bob

Helen (pseudonym), using a one-way analysis of variance to test if the mean instances of on-task observations differed significantly at the *a<.05* level between phases, resulted in a finding that no significant differences existed in the means of on-task data between any phase condition [*F(2,21)=1.81, p=.188*]. Within the planned contrast examinations, no significant affect was shown between baseline and the TE implementation (both physical and virtual combined) [$t(19)=1.718$, $p<.05$] (one tailed) nor was any difference between physical and virtual implementation methods found [$t(19)=.102$, $p<.05$] (two tailed). Note that the examination of contrast between the means of virtual and physical implementation phases combined with regard to baseline indicated a one-tailed significance of $p=.051$. While this was not strictly significant statistically, it is worth noting that this test barely missed the levels required. Overall single case plot visual analysis did not indicate any discernable patterns across implementation phases (see Figure 3). The single case visual analysis provided further cause to support a finding of non-significant in regard to the virtual vs physical implementation planned contrast examination that was so close to a rounded $p=.05$ cutoff point.

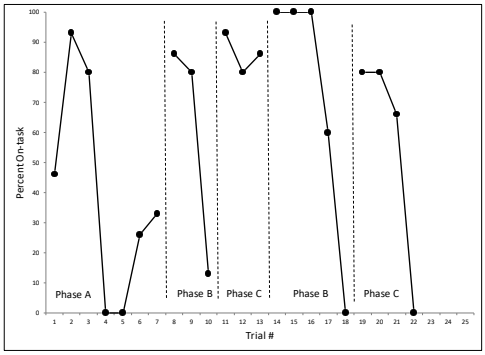

Figure 3. Helen

Mark (see Figure 4) showed an overall decrease in time on task over baseline achievements. A one-way analysis of variance was conducted to test if the mean instances of on-task observations differed significantly at the *p<.05* level. Results indicated that no significant differences existed in the on-task instances data between any condition [*F(2,19)=1.122, p*<.05]. Further analysis using planned contrasts indicated that no significant affect was shown between baseline and the TE implementation (both physical and virtual combined) [$t(19)= -1.415$, $p<.05$] (one tailed) nor was any difference between physical and virtual implementation methods found [$t(19)=-.545$, $p<.05$] (two tailed). Note that because Mark's visual mean plot data indicated a negative slope, the planned contrast concerning differences in the combined TE methods and baseline at the two tailed level were also not significant at $p<.05$. Overall visual analysis of single case data did not indicate any observable trends in data between phases of TE implementation (see Figure 4).

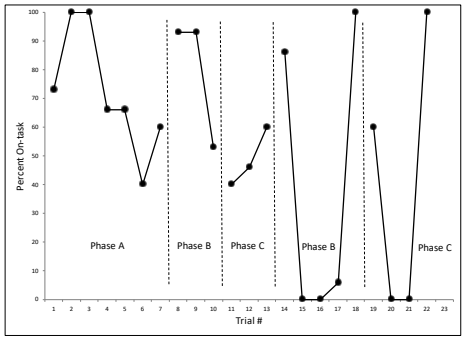

Figure 4. Mark

4.1. Social Validity

Social validity was obtained via a student and teacher questionnaire given to the students following the data collection periods. Student questionnaires contained three questions. First, students were asked if they preferred the paper or iPad token delivery system. Twenty students responded to this question and 16 indicated a preference for the iPad delivery. One student stated that "…it was cool to see the points pop up", while another noted that the iPad was preferred because "…you don't have to count the points". Two

students preferred the paper token delivery stating that such system allowed them to "…share it (paper tokens) with Friends [SIC] if you want to save up for something like Raptor room." Two students stated that they had equal preference for paper or iPad-based tokens.

Question two asked students to rate, on a scale of one to ten, how much they focused on obtaining tokens through being 'on task' during periods of using either of the two systems. Results indicated that 7 of 16 (43.8%) rated their attention to obtaining tokens as a 1 (did not focus on obtaining tokens at all). One student stated, "After a while, the thought of getting a reward wore down". Another three students (18.7%) gave the score of 3 and three more gave a score of 5. One of the students that indicated a 3 stated that they only focused on being on task to obtain tokens about a quarter of the time "…because your [SIC] so busy working." Two students indicated a 10 in response to question 2 and stated that they focused on receiving tokens "…all the time."

Asked in question three, which method (iPad, Physical, Both, None) they would recommend teachers use to help students focus on their work, one indicated paper, five indicated both and ten indicated iPad. One student that had indicated that they would recommend both systems to teachers stated that they did so "…cuz [SIC] then there would be two ways of getting rich!" One student that indicated they would recommend the iPad method stated they did so "… because it (tokens) can't be stolen." and "Because its [SIC] fun". It should be noted that early in the implementation, one instance of theft of physical tokens (bills) occurred (and was rectified by the teacher). This likely directly related to this student's reference to such possible issues on the anonymous survey.

The teacher participant also provided social validity feedback data through a separate questionnaire. Overall, the teacher participant indicated that the paper methodology was more effective in helping keep students on task. The teacher indicated that the paper method provided "…instant gratification… students knew why they earned the token… It caused a ripple effect around the student who earned the token, that others (would) see what happened and learn that if they did the same thing, they too could earn a token." As a corollary, the teacher stated that "…(using) the iPad system, students did not see when someone (else) earned a token because it only showed up on the individual who earned the (token on their) iPad." Further, the teacher noted the iPad app was difficult and time consuming to use.

5. Discussion

It is interesting to note that prior to the results of the present research being presented to the subject participants, the teacher indicated an overall satisfaction with the TE as a classroom management method. The teacher indicated that she felt the overall attention to task for students increased during times in which she implemented the TE methods. The results seemed to be surprising to the teacher when reveled at a classroom pizza party following the study.

5.1. Token delivery

A variable ratio schedule of reinforcement (token delivery) is generally accepted as effective regarding the tracking and reinforcement of on task behaviours [17, 18]. The teacher in the present study also requested this reinforcement schedule so that time to deliver tokens, both virtual and physical, could occur when breaks in her teaching flow allowed and so that instruction would not be interrupted based on a fixed interval reinforcement methodology. One possible explanation for the overall ineffectiveness of both token economy systems in the present study may be related to the variable ratio schedule of reinforcement. The reinforcement schedule that resulted from relying on breaks in lesson flow may have been sub-optimal for some students.

It is therefore possible that prior to the implementation of a variable ratio reinforcement schedule, students may require a more defined schedule of interval reinforcement prior to the application of a variable ratio methodology. Future researchers should consider this possibility as well as the equally possible reality that such alterations in delivery schedule may be impractical for a teacher to administer alone. Further study is required to address such hypothesis.

Another area of interest was the non-public nature of token delivery during the iPad based TE phases. It may be that when students noticed delivery of tokens, they made an effort to display the desired on-task behaviour but the behaviour might have dissipated when students noticed the teacher otherwise engaged. If this had been the case, we likely would have expected to see a difference in impact between the private iPad deliver and the public physical deliver of tokens. This was not the case in the present study.

Despite the teacher's best intentions, within the current study framework, she was unable to attend to the on-task behaviour of the group 100% of the time while teaching either group or station-based lessons. This would seem to indicate that the need to physically deliver tokens versus being able to do so from a distance did not impact the teacher's ability to attend

to the behaviours for which tokens were to be delivered. The teacher seemed to confirm this suspicion by stating in the follow-up questionnaire that "…allow(ing) the EA (educational assistant) to hand out the tokens instead of the teacher" for paper delivery would be helpful and simplifying the finding of specific students within the app's interphase would reduce the difficulty in delivering tokens to individuals and/or small groups of individuals. These assertions by the teacher seem to indicate difficulty with being able to teach while simultaneously attending to the observation of student on-task behaviours.

As with the previous assertion concerning public vs private token delivery, physically walking over to students (physical) vs taping an iPad (iPad) to deliver tokens did not seem to impact the effectiveness of either methodology as to impact on student on task behaviour. We would have expected to see a difference between the impact between physical and iPad-based methods if delivery method had been an important aspect of the method however this was not observed.

It is likely that the need to simultaneously focus on the fluid needs of instruction while teaching allows for limited attention to matters of observation regarding individual or group behaviours. Indeed, the teacher's token delivery occurred during times within lessons that did not require her direct involvement with a student. This hypothesis would seem to support recent research regarding a teacher's ability to mulit-task. As cognitive tasks are divided between two or more pressing needs, the quality and efficiency of results is generally reduced [19, 20, 21]. Such a finding regarding teacher abilities to multi-task would seem to point to one possible reason for the overall failure of the TE system in the present study.

5.3. Token redemption

Token redemption took place at least one time per day at one or more pre-determined redemption periods, however the students were required to ask the teacher for redemption during the noted times. Sometimes the teacher was otherwise engaged during these times, speaking with other faculty members while children were at play or preparing stations for when children would return. Occasionally, the teacher was required to serve as a recess monitor and was unavailable to deliver tokens during recess. Overall, this resulted in a less predictable token redemption time period during both phases of TE implementation. Students may have been discouraged if they had intended on receiving a prize at a specific time period in which the teacher was unable to comply with a purchase request. While students had been told that not all the redemption periods would be available due to the teacher's multiple commitments, and that one would be available at minimum per day, the lack of a solid, repetitive daily redemption schedule may have negatively impacted the students' motivation to remain on task.

Again, future researchers should address this redemption hypothesis in more detail to examine any impact a more predictable redemption schedule may have upon the overall time on task behaviours of students. Like the delivery hypothesis, researchers must also seek to understand if a predictable redemption schedule is reasonable to maintain when the teacher alone, implements the TE system. It may be the case that additional help may be required if predictable delivery of tokens and predictable redemption periods other than the one time per day in the present study are to be achieved.

5.4. Analysis of efficacy

Results indicated that the virtual delivery TE system and the combined data from virtual and physical methods were significantly effective over baseline (no TE) for Bob only. No other individual or whole group analysis showed a significant difference between base line and the two TE approaches nor between the two TE approaches themselves. This may indicate that in spite of statistical indications, the delivery of tokens to Bob was optimal or effective by sheer chance alone (within the 5% error range). Also, Bob's data included an outlier in data point five (score of 0). No obvious reason for Bob's inattention during that data observation period was noted and thus for official analysis, the point remained within the data set. It is important to note, however, that this possible outlier influenced the magnitude of significant results. Adding visual assessment of raw data, it seems that at best, we can describe the results for Bob as inconclusive.

While the current findings indicated support for the findings of Maggan et al., and Ivy et al., [12, 2], the current work would seem to contradict some other available research regarding the effectiveness of TE systems within an inclusive classroom setting. Given the negative results of the present work in relation to previous studies suggest that clarity in the implementation of studied TEs is critical to understanding conclusions drawn from any findings. In the present work, implementation fidelity was strictly noted and adhered to a pre-defined set of standards. Given those standards, results showed the method as implemented not to be an effective support regarding on task behaviours within the student population studied. When one considers the incredible differences with which the idea of a TE can be implemented (ie: multiple human intervention agents,

the behaviour/s of focus, the diversity of students individual characteristics, the token delivery and redemption schedules), it is likely not possible to assert that any 'generic' TE method should be the focus for analysis leading to the categorization of evidence based practice. Instead, specific versions of the TE, strictly defined, may be a more proper unit of analysis.

6. Limitations

This work is limited to that observed within the contexts of the participants within the location chosen for the study. Results should not be used to justify broader meaning outside of this context, as individual circumstances exist in any defined population and context of study.

Additionally, time on task represents a difficult variable of measure. Specifically, data collectors were required to identify the direction of each subject's attention or activity toward a direction, activity or object that was relevant to the instruction being provided at that time while simultaneously excluding indicators of non-attention to task as defined by Lee, Sugai and Horner [15]. The relevance of the direction of attention or activity based on instruction can be somewhat subjective to the person judging the data point. For example, if a student is looking at his/her shoes while the teacher is working mathematics on a white board, the data recorder would likely mark the data point as 'not on task' however if the teacher were using eyelets of shoes as an example to count pairs of objects, the same gaze would be recorded as 'on task'. IRR was used to indicate the breadth of subjectivity with reasonable findings however it is important to acknowledge such as a limitation to the results of the present work.

7. Acknowledgements

This research was supported by the Social Sciences and Humanities Research Council of Canada.

8. References

[1] Hackenberg, T.D. (2018). Token reinforcement: Translational Research and Application. *Journal of Applied Behavior Analysis, 51,* 393-435. doi: 10.1002/jaba.439

[2] Ivy, J.W., Meindl, J.N., Overley, E. & Robson, K.M. (2017). Token Economy: A Systematic Review of Procedural Descriptions, *Behavior Modification, 00*(0), 1-30. doi: 10.1177/0145445517699559

[3] Kazdin, A.E. (1977). *The token economy: A review and evaluation,* New York: Plenum Press.

[4] Robacker, C.M., Rivera, C.J. & Warren, S.H. (2016). A Token Economy Made Easy Through ClassDojo, *Intervention in School and Clinic, 52*(1), 39-43. doi: 10.1177/1053451216630279

[5] Simonsen, B., Fairbanks, S., Briesch, A., Myers, D., & Sugai, G. (2008). Evidenced-based practices in classroom management: Considerations for research to practice. *Education and Treatment of Children*, 31. 351–380.

[6] Carnett, A., Rulston, R., Lang, R., Tostanoski, A., Lee, A., Sigafoos, J. & Machalicek, W. (2014). Effects of a Perseverative Interest-Based Token Economy on Challenging and On-Task Behavior in a Child with Autism, *Journal of Behavior Education, 23,* 368-377. doi: 10.1007/s10864-014-9195-7

[7] Soares, D.A., Harrison, J.R., Vannest, K.J. & McClelland, S.S. (2016). Effect Size for Token Economy Use in Contemporary Classroom Settings: A Meta-Analysis of Single-Case Research, *School Psychology Review, 45*(4), 379-399

[8] Kazdin, A.E. (1982). The token economy: A decade later., *Journal of Applied Behavior Analysis, 15,* 431-445.

[9] Matson, J.L. & Boisjoli, J.A. (2009). The token economy for children with intellectual disability and/or autism: A review, *Research in Developmental Disabilities, 30,* 240-248. doi: 10.1016/j.ridd.2008.04.001

[10] Alter, P. (2012). Helping Students with Emotional and Behavioral Disorders Solve Mathematics Word Problems, *Preventing School Failure, 56*(1), 55-64, doi: 10.1080/1045988X.2011.565283

[11] Ogasahara, K., Hirono, M. & Kato, S. (2013). Support for on-task behavior through a token economy system: Autistic youth who shows challenging behavior, *Jamanese Journal of Special Education, 51*(1), 41-49

[12] Maggin, D. M., Chafouleas, S. M., Goddard, K. M., & Johnson, A. H. (2011). A systematic evaluation of token economies as a classroom management tool for students with challenging behavior. *Journal of School Psychology*, 49, 529-554.

[13] Hackenberg, T.D. (2009). Token reinforcement: A review and analysis. *Journal of the Experimental Analysis of Behavior, 91,* 257-286. doi: 10.1901/jeab.2009.91-257

[14] Kazdin, A.E., & Bootzin, R.R. (1972). The token economy: An evaluative review. *Journal of Applied Behavior Analysis, 5,* 343-372.

[15] Lee, Y.Y., Sugai, G., & Horner, R.H. (1999). Using an instructional intervention to reduce problem and off-task behaviors. *Journal of Positive Behavior Interventions, 1*(4), 195-204

[16] Alberto, P. & Troutman, A. (2003). *Applied Behavior Analysis for Teachers: 6th Ed.*, Merrill Prentice Hall, Upper Saddle River, New Jersey, Columbus Ohio

[17] Martens, L.K., Lochner, D.G., & Kelley, S.Q. (1992). The effects of variable-interval reinforcement on academic engagement: A demonstration of matching theory. *Journal of Applied Behavior Analysis, 25,* 143-151

[18] Hulac, D., Benson, N., Nesmith, M.C. & Shervey, S.W. (2016). Using Variable Interval Reinforcement Schedules to Support Students in the Classroom: An Introduction With Illustrative Examples, *Journal of Educational Research and Practice, 6*(1), 90-96, doi:10.5590/JERAP.2016.06.1.06

[19] Bowman, L.L., Levine, L.E., Waite, B.M., & Dendron, L. (2010). Can students really multitask? An experimental study of instant messaging while reading. *Computers & Education, 54*(4), 927-931, doi: 10.1016/j.compedu.2009.09.024

[20] Logie, R.H., Gilbooly, K.J., & Wynn, V. (1994). Counting on working memory in arithmetic problem solving. *Memory & Cognition, 22*(4), 395-410, doi: 10.3758/BF03200866

[21] Wieth, M.B. & Burns, B.D. (2014). Rewarding Multitasking Negative Effects of an Incentive on Problem Solving Under Divided Attention, *Journal of Problem Solving, 7*, doi: 10.7771/1932-6246.1163