Memo

To:	Office of Information and Regulatory Affairs (OIRA), Office of Management and Budget (OMB)
From:	Marie Lawrence, Administration for Children and Families (ACF), Department of Health and Human Services (HHS)
Date:	6/9/2025
Subject:	NextGen Follow-up Survey Gift Card Experiment

The Mathematica study team conducting the Next Generation of Enhanced Employment Strategies (NextGen) Project, under contract to the Office of Planning, Research, and Evaluation (OPRE) within the Administration for Children and Families (ACF), is continuing to identify strategies to boost response rates on follow-up surveys of study participants. Starting in June 2022, the study team provided a \$5 prepaid cash token of appreciation to some participants before they responded to the follow-up survey as part of an experimental test to improve the response rate. We have now analyzed the findings and would like to provide an update on the follow-up survey prepay experiment, along with recommendations for next steps. **Our findings indicate that the \$5 pre-paid cash did not have an impact on overall response rates.** This memo presents background on the experiment and its implementation, findings from the analysis, a discussion of the findings, and a list of recommended next steps.

1. Motivation for conducting an experiment with the follow-up survey gift cards

The purpose of the NextGen Project is to test the effectiveness of innovative, promising employment interventions designed to help individuals with low incomes facing complex challenges secure a pathway toward economic independence. The project is conducting four randomized controlled trials, and data to measure the interventions' effects on employment, earnings, and other outcomes of interest are collected, in part, through follow-up surveys with study participants.

Individuals participating in the NextGen Project are members of groups that are traditionally considered "hard-to-reach." These populations include people with mental illness, young adults with disabilities, individuals who have been recently incarcerated, those with very low incomes, and those with combinations of these factors. The risk of biased impact estimates increases with lower overall survey response rates or larger differences in survey response rates between key research groups (What Works Clearinghouse 2022). Continued high rates of participation in the study, through the second follow-up survey, are necessary to produce unbiased estimates of the program impacts and maximize the utility of survey data in this multipart study.

Tokens of appreciation, such as gift cards, are a long-established, effective method of increasing survey participation and support retention over the course of a longitudinal data collection. In the early stages of the project, the study team was concerned about response rates given the hard-to-reach populations and was looking for token of appreciation designs that could mitigate some of the anticipated challenges. A prepay token design is one promising way to overcome the challenges, given their success with general populations, as discussed below. As described in the project's Office of Management and Budget (OMB) clearance requests (OMB Number 0970-0545), the NextGen Project team decided to conduct an experimental test of a \$5 prepaid cash token of appreciation offered before a sample member responds to the first follow-up survey. This experiment allowed Mathematica to test the effectiveness of one approach

Date: 6/9/2025 Page: 2

for increasing the response rates. Prepaid tokens of appreciation may be effective because they lend credibility to a postpaid token and to the project more generally. In addition, they can create a sense of reciprocity. Behavioral science literature suggests that the norm of reciprocity requires that we repay in kind what another has done for us; in this context, giving a potential survey respondent five dollars should increase their desire to reciprocate by completing the survey (Falk 2007; Gneezy and List 2006; Cialdini 2007).

Studies across general populations show prepaid tokens of appreciation to be more effective than postpaid ones, and the combination of prepaid and postpaid is more effective than prepaid alone (for example, the systematic review in Singer and Ye 2013 and the meta-analysis of household surveys in Mercer et al. 2015). In a meta-analysis of 39 experimental studies, Singer et al. 1999 showed that a prepaid token can yield a higher response rate compared to a postpaid scheme for interviewer-administered surveys. Cantor et al. (2008) found that prepaid tokens of appreciation between \$1 and \$5 increase response rates from 2 to 12 percent compared to no prepaid or postpaid tokens of appreciation for random digit dial telephone surveys. Mercer et al. (2015) conducted a meta-analysis of published and unpublished experiments since 1987 and found that prepaid tokens of appreciation are more effective than postpaid ones. Jäckle and Lynn (2008) found that prepaid tokens of appreciation significantly reduced attrition in follow-up waves under a longitudinal design. Hock et al. (2015) found that a \$5 prepaid token of appreciation for a hard-to-reach population (unemployed adults) yielded a higher response rate for online mode in a phone and web administered survey.

Although the survey and behavioral science literature shows strong support for the effectiveness of prepaid tokens of appreciation at increasing overall response rates with general populations, less is known about their usefulness specifically in surveys of populations with low incomes, or those considered "hard-to-reach," participating in randomized control trials of employment programs. We designed the NextGen experiment to address this knowledge gap. In addition, we used the experiment as an opportunity to test if a prepaid token has any impact on the study's program and comparison response rate differential. For the experiment, one research group (treatment) was offered a \$5 prepaid token and a \$50 postpaid token, and the other research group (control) was offered no prepaid token of appreciation but offered a \$55 postpaid token of appreciation.

The main questions we intended to answer with this experiment were:

- Does a prepaid token of appreciation, combined with a postpaid one at survey completion, increase response rates compared to only the postpaid one?
- Does a prepaid token of appreciation, combined with a postpaid one at survey completion, decrease response rate differentials across study groups, relative to only the postpaid one?

2. Implementation of the experiment

The experiment randomized all NextGen Project participants into two groups for the prepaid token experiment: prepay (treatment) group and no prepay (control) group. We used a simple randomization scheme that targeted a 50-50 split, conducted just prior to the launch of the survey effort within each survey cohort. Both groups received the same total dollar amount, \$55. The prepay group received a \$5

Date: 6/9/2025 Page: 3

bill attached to the survey's advance letter and received a \$50 gift card upon completion of the survey. The no prepay group only received a \$55 gift card upon survey completion.

The randomization for the experiment began in June 2022 when we invited the first cohort of NextGen participants to participate in the first follow-up survey. Since that time, we randomized all subsequent cohorts prior to their survey invitations.

To ensure the soundness of the experimental design, Mathematica conducted thorough quality assurance reviews during the experiment. These included:

- Reviewing all advance letter mailings, prepay and no prepay, to ensure we sent the correct version of the letter, the name on the letter matched the name and address on the envelope, the \$5 bill was included in prepay group letters, and the \$5 bill was not included in the no prepay group letters.
- Testing all survey systems to ensure that all subsequent communications (text messages, emails, interviewer outreach, reminder mailings) match the experiment's post-pay amount (\$50 or \$55).
- Reviewing all payment mailings, prepay and no prepay, to ensure we included the correct postpay amount for each individual.
- Working with field locators to ensure they understood which participants are part of which experiment group based on the indicator in the field management system and training the field locators to ensure they only discuss the correct post-pay amount with each participant.
- Reviewing raw experimental outcomes periodically to ensure that we enforced the experimental groups correctly and that the experiment was not causing large response rate discrepancies that could affect the NextGen Project's impact study.

Since the launch of the survey in June 2022 through October 2024, we randomized 2,779 cases into the experiment, with 1,356 in the prepay group and 1,423 in the no prepay group. In the analysis that follows, we excluded two batches of cases. First, we excluded 15 cases from the Families Achieving Success Today (FAST) program¹. Second, we excluded all 46 cases from the fifth cohort (all survey cases with a survey launch during July 2022). Our team did not conduct the full set of quality assurance steps for the advance mailing for the fifth cohort, so we cannot guarantee that all prepay cases correctly received the \$5 prepay or that all those in the no prepay group cases did not.

The final number of cases with these exclusions is 2,718 cases (1,331 in the prepay group and 1,387 in the no prepay group). In analyses below that use baseline characteristics, a further 64 cases are excluded as they do not have a full set of baseline characteristics. For these analyses, there are a total of 2,654 participants (1,309 in the prepay group and 1,345 in the no prepay group).

We constructed baseline characteristics using respondent-reported information from the baseline survey for variables we thought would impact the prepay response rate. They include sex, race, education, language, and unstable housing.

• Sex. Based on the intake question asked during study enrollment.

¹ FAST was an early NextGen site, but the program dropped from the project, and we ended all follow-up survey efforts. At that time, we had only randomized 15 cases into the experiment and sent invitations to take the survey. Since we did not complete the survey outreach protocol for these cases, we excluded them.

Date: 6/9/2025 Page: 4

- Race. An indicator for if respondent selected non-Hispanic ethnicity at item B1 and Black or African American at item B2 (race) on the baseline survey. Respondents could select multiple options.
- Education. An indicator for education level of some college or more, based on item B5.
- Language: An indicator of if Spanish is spoken at home, based on item B3.
- Unstable housing. An indicator defined if housing status in the past month was homeless or in emergency housing such as a shelter, a halfway house, sober house or other transitional housing, a group home, or living with friends or relatives and not paying rent., based on items B9 and B9a of the baseline survey.

We reviewed the baseline characteristics by pre-pay experiment groups to check if the experiment's randomization scheme led to any significant, unexpected differences. None of the baseline characteristics show any significant differences between the prepay and no prepay groups (Table 1), indicating that there are no systematic issues with the experiment's randomization process.

Baseline		Proportion of sample	Proportion of sample		
characteristic	Sample size	(no prepay)	(prepay)	Test statistic	p-value
Sex (female)	2,716	64.4	63.3	-1.1	0.562
Race (non- Hispanic black)	2,694	41.6	40.0	-1.6	0.408
Education (some college or more)	2,713	29.4	29.7	0.3	0.845
Language (Spanish)	2,718	13.3	14.3	1.0	0.445
Unstable housing	2,682	27.1	26.9	-0.2	0.924

Table 1. Baseline characteristics by prepay token experiment group

***/**/* The differences were statistically significant at the 1, 5, and 10 percent levels, respectively, using a two-tailed t-test.

3. Prepay token findings

Overall findings and within NextGen Programs

We find that the prepay and no prepay groups had similar response rates and the differences were not statistically significant at the 0.10 level (Table 2). Results are similar when adjusted for the set of baseline characteristics: sex, race, education, language, and unstable housing.

Table 2. Overall impact of the prepay token on survey response rate

Model	Sample size	Response rate (no prepay)	Response rate (prepay)	Impact of prepay	p-value
Unadjusted	2,718	70.4	67.8	-2.7	0.132
Adjusted for baseline	2,654	70.4	67.7	-2.8	0.119

Date: 6/9/2025

Page: 5

Model	Sample size	Response rate (no prepay)	Response rate (prepay)	Impact of prepay	p-value
characteristics					

***/**/* The differences were statistically significant at the 1, 5, and 10 percent levels, respectively, using a two-tailed t-test.

We analyzed the results by each NextGen program adjusting for baseline characteristics to see if the response rates were different by experimental group (Table 3). There were no statistically significant differences in response rates by experimental group for three out of four programs; for one program, Individual Placement and Support for Adults with Justice Involvement (IPS-AJI), a negative impact emerged.

Table 3. Impact of the prepay token on survey response rate by NextGen program

	Comple cize	Response rate	Response	Impact of	n velue
NexiGen Program	Sample Size	(no prepay)	rale (prepay)	prepay	p-value
Bridges from School to Work (Bridges)	690	69.4	71.5	2.1	0.548
Individual Placement and Support for Adults with Justice Involvement (IPS-AJI)	601	61.6	52.7	-8.8**	0.029
Western Mass Mental Health Outreach for Mothers (MOMS)	716	82.2	83.1	0.9	0.746
Philadelphia Workforce Inclusion Networks (Philly WINS)	647	66.8	60.7	-6.1	0.106

Note: Adjusted for baseline characteristics.

***/**/* The estimated differences were statistically significant at the 1, 5, and 10 percent levels, respectively, using a two-tailed t-test.

Subgroup analysis

We also conducted subgroup analyses for the survey response rate using each of the baseline characteristic variables within the prepay experiment groups. We conducted these analyses to further understand what might be driving the no-impact (or negative impact from IPS-AJI) findings. If the prepay experiment has a different effect on subgroups of participants, it could potentially lead to seeing an unexpected impact like we observed with IPS-AJI. For each NextGen program, we examined whether impacts on the response rates differed for subgroups according to the study participant's (1) race, (2) language, (3) unstable housing, (4) education, and (5) sex. Only one subgroup pairing had statistically significant differences in impacts, which is unstable housing for overall sample and for IPS-AJI. Table 4 shows that among NextGen participants who were unstably housed at baseline, the prepaid gift card had a significant negative impact on response rates of -8.0 percentage points. Among people who were unstably housed and participating in the IPS-AJI program, the prepay gift card had a significant, negative impact

Date: 6/9/2025 Page: 6

of -11.2 percentage points. In the other programs, the prepaid gift card does not show a statistically significant impact.

NextGen Program	Response rate (no prepay)	Response rate (prepay)	Impact of prepay	p-value
All NextGen programs	65.3	57.3	-8.0**	0.026
Bridges	66.7	54.6	-12.1	0.614
IPS-AJI	62.6	51.3	-11.2**	0.027
Western Mass MOMS	81.8	74.7	7.1	0.346
Philly WINS	59.1	55.3	-3.8	0.605

Table 4. Unstable housing impact on prepay response rates by NextGen program

Note: Adjusted for baseline characteristics.

***/**/* The estimated differences were statistically significant at the 1, 5, and 10 percent levels, respectively, using a two-tailed t-test.

As discussed above, unstable housing is a baseline indicator that flags participants that were homeless or in emergency housing such as a shelter, a halfway house, sober house or other transitional housing, a group home, or living with friends or relatives and not paying rent. We attempted to conduct a robustness check by defining unstable housing using other data that was available, namely follow-up survey paradata². We pulled paradata from the survey effort at the participant level that included indicators if a case ever had an address in the survey database marked as bad, the case ever had a phone number marked as bad, or the case ever had an address identified as "homeless."

We find consistent patterns using this different data source for unstable housing: unstable housing seems to play a role in the negative effect of the prepay token for the IPS-AJI program. However, the results are not as strong as the baseline characteristic variable, likely due to the quality of the survey paradata.³ The IPS-AJI program also has a much higher proportion of participants experiencing unstable housing (63.4 percent) compared to the other programs (an average of 16.5 percent for Bridges, MOMS, and Philly WINS). The survey paradata shows similar differences. For instance, 46.1 percent of IPS-AJI cases show at least one address marked bad, compared to under 21 percent for each of the other three sites.

Differential response rates

In addition to exploring the effect of the prepay on overall response rates, we explored whether the prepay had an effect on the treatment-control differential response rate. There were no statistically significant differences in response rates by prepay and no prepay. Within the prepay group, the response rate

² Survey paradata is data collected as part of the survey process itself. It includes data like minutes to complete, address history records, number of telephone calls to a case and similar data.

³ The quality of this survey paradata is mixed since it is not always available and does not apply to all cases equally. As an example, if the Post Office returns mail as undeliverable, we will mark that address as bad in our survey database. However, the Post Office does not always return mail as undeliverable. In this situation, an address would only be marked as bad if our address check software flags the address as bad, or a field interviewer discovers it is bad. If the participant completes the survey before this happens (such as completing the web survey in response to an email or text message invite), the bad address will never be discovered as part of the survey fielding.

Date: 6/9/2025 Page: 7

difference between program and comparison was 1.1 points; this gap was also 1.1 points within the no prepay group.

4. Discussion and conclusion

Overall, the prepay token did not impact the response rate. This is true at the project level (i.e., across all NextGen study participants), but not at the program level because in IPS-AJI we find a significant negative impact of the prepay token (those who receive the \$5 prepay are less likely to complete the survey). We find these results surprising and go against the initial hypothesis of the experimental design. Based on the available evidence, we suspect that instability in housing is a contributing factor to this finding. The available data is not sufficient to prove a relationship, but we propose several anecdotal theories to explain it.

The first idea is an effect based on communication between NextGen participants, which could work in multiple ways. Consider a participant offered a \$50 post-pay to complete the survey but does not recall receiving the \$5 prepay or never received the prepay due to a bad address. If that participant hears that another participant is offered \$55 to complete the survey, it may discourage participation. Similarly, if they do not recall or did not receive a prepay and then hear that another respondent received a \$5 prepay to complete the survey, even if both are offered the \$50 post-pay, it may also discourage participation. We believe that this type of communication between participants is more likely in IPS-AJI compared to other programs. IPS-AJI participants are more likely to find themselves in circumstances that place them in close quarters with each other: group homes, food pantries, shelters, and homeless encampments which are, anecdotally, more prevalent among participants of IPS-AJI. IPS-AJI also has a higher proportion of participants with unstable housing according to the baseline data.

A second idea is an effect based on token amount. A \$55 post-pay amount may encourage more participation that \$50 post-pay and this effect is magnified with unstable housing conditions where a larger proportion of \$5 prepay did not successfully reach the participant. The NextGen survey communication protocols do not refer to the \$5 prepay token after the initial letter. So, if the advance letter does not reach the participant, any \$5 prepay participants do not know they were supposed to get one and only hear communications about a \$50 post-pay token.

In hindsight, our design for the experiment may not have been ideal based on our choice of envelopes. More recent literature suggests that pre-pay efforts work best with visible cash prepay tokens (DeBell 2023). This involves using an envelope with a window that allows the sample member to see the token of appreciation, which makes it more likely they will open the envelope. We considered this when we were discussing the initial design, but since these were low-income populations, we worried that the money would be removed by others before ever reaching the sample member. By choosing standard envelopes that did not include any indicator that there was money inside, we may have removed an enticement to open the letter. In future experiments, it would be interesting to test this theory.

Another issue that might explain why the experiment didn't work is that prepay tokens work best with populations that are new to an evaluation (Singer and Ye 2013). They are a good tool for recruiting efforts because they legitimize the study by offering money upfront and suggest to people that it's not a scam and they will receive the full token of appreciation. Since our sample had already been recruited, we likely did not receive the same benefit from reassuring our legitimacy that most recruiting methods did. It

Page: 8

is possible that the results of prepay tokens of appreciation in longitudinal studies may not be as positive as they are for recruiting.

5. Next steps

The first follow-up survey is nearing the end of its fielding period. The final cohort launched in March 2025 and Mathematica will likely continue the data collection until mid-2025. Given that the prepay token does not show an impact on the response rate overall and negatively impacts the response rate for one NextGen program, we ended the experiment. We will not implement any changes for the second follow-up survey since the pre-pay token did not prove successful. For the second follow-up survey, we will provide no pre-pay tokens and \$50 electronic gift cards following survey completion to all study participants.

References

- Cantor, David, Barbara O'Hare, and Kathleen O'Connor. 2008. The use of monetary incentives to reduce non-response in random digit dial telephone surveys. In Advances in telephone survey methodology, eds. James M. Lepkowski, Clyde Tucker, J. Michael Brick, Edith de Leeuw, Lilli Japec, Paul J. Lavrakas, Michael W. Link, and Roberta L. Sangster, 471–98. New York, NY: Wiley.
- Cialdini, R. B. (2007). Influence: the psychology of persuasion. Rev. ed.; 1st Collins business essentials ed. New York: Collins.
- DeBell, Matthew. "The Visible Cash Effect with Prepaid Incentives: Evidence for Data Quality, Response Rates, Generalizability, and Cost." Journal of Survey Statistics and Methodology 11.5 (2023): 991-1010.
- Falk, Armin. "Gift Exchange in the Field." *Econometrica* 75, no. 5 (2007): 1501–11. http://www.jstor.org/stable/4502037.
- Gneezy, Uri, and John A. List. "Putting Behavioral Economics to Work: Testing for Gift Exchange in Labor Markets Using Field Experiments." *Econometrica* 74, 5 (September 2006): 1365-1384.
- Hock, Heinrich, Priyanka Anand, Linda Mendenko, Rebecca DiGiuseppe, and Ryan McInerney. The Effectiveness of Prepaid Incentives in a Mixed-Mode Survey. Presentation to the Annual Conference of the American Association for Public Opinion Research 2015.
- Jäckle, Annette, and Peter Lynn. 2008. Respondent incentives in a multi-mode panel survey: Cumulative effects on nonresponse and bias. Survey Methodology 34 (1): 105–17.
- Mercer, Andrew; A. Caporaso; D. Cantor; and R. Townsend. 2015. How Much Gets You How Much? Monetary Incentives and Response Rates in Household Surveys. *Public Opinion Quarterly*, Vol. 79, No. 1, Spring 2015, pp. 105-129.
- Singer, Eleanor, and Cong Ye. 2013. "The Use and Effects of Incentives in Surveys." ANNALS of the American Academy of Political and Social Science 645:112–41.
- Singer, Eleanor, Nancy Gebler, Trivellore Raghunathan, John Van Hoewyk, and Katherine McGonagle. 1999. The effect of incentives in interviewer-mediated surveys. *Journal of Official Statistics* 15 (2): 217–30.

Date: 6/9/2025

Page: 9

What Works Clearinghouse. "Standards Handbook, Version 5." 2022. Available at <u>https://ies.ed.gov/ncee/WWC/Docs/referenceresources/Final_WWC-HandbookVer5_0-0-508.pdf</u>.