Use of tax terminology and deadline language in tax benefits outreach

Date implemented: November 2022
Date written: January 2023
Abstract: This study reports the results of an experiment varying the language of tax benefits outreach. On November 9, we messaged 236,476 clients who had either used GetCTC in 2021 or had signed up for updates about the service earlier in 2022, encouraging them to use GetCTC. Half of clients received a message using non-tax language ("cash benefits"), while the other half saw tax language ("tax credits"). Meanwhile, half received a message mentioning the impending November 15 deadline, while half saw a message with no deadline. We find that non-tax language and deadline language significantly increase home page visits, likely increase the number of returns started, and may increase the number of returns submitted—but neither has a detectable impact on the number of returns accepted. The results suggest that, on the true outcome of interest, the outreach wording is not a significant consideration, and they also highlight the danger of analogous studies that look only at top-of-the-funnel outcomes.
Authors: Gabriel Zucker, Maximilian Hell

Other experimental results and research from GetCTC 2022 are available here.

---

1. Research Questions

This study addressed two questions regarding the messaging of tax benefits outreach.

First, the study investigated the impact of using explicit tax language in tax benefits outreach. Research from 2021 (see Lessons from Simplified Filing 2021, p. 80) and a study using GetYourRefund in April 2022 both suggested that using explicit tax language (i.e., tax benefits, credits) underperformed more generic benefits language (i.e., payments, cash benefits) among traditional non-filers. This study sought to replicate those findings.

Second, the study investigated the impact of including language around deadlines in tax benefits outreach. November 15 was the last advertised day for new clients to start returns on GetCTC (though technically there was a 24-hour grace period, through November 16), but many outreach partners had opted not to include notice of this deadline in their final messages. We hypothesized that mentioning the deadline might help focus clients who might otherwise not feel the urgency to start their return, leading to an increase in filing. The study was intended to test this hypothesis.

2. Study Design and Implementation

The study contained two distinct cohorts. Cohort 1 contained successful GetCTC users from 2021 who had entered at least one dependent in 2021, had opted in to receiving text messages, and who had not yet started GetCTC returns in 2022. This cohort had received up to four earlier rounds of messaging. Within this cohort, only English speakers were included in the experiment; Spanish speakers received messaging, but it was not randomized and is not included in the study. Cohort 2 contained clients who arrived at GetCTC.org prior to May 11, 2022, had signed up to receive updates about the service's availability, and had not started an intake using the same contact method before November 9. (These clients' Social Security Numbers are not known, so they can only be matched to existing returns via the contact method they used to sign up for updates.) This cohort had received one prior round of messaging, in late May. Clients were only included in the study universe if they had signed up for updates via text message. Those who signed up via email received a message but were not included in the study.

The study was implemented as a matrix design, with four possible scripts, with two variants based on the cohort. The scripts and the number of recipients in each cohort are shown below:

		Tax	No tax
Cohort 1	Deadline	22,845 recipients Hello! This is GetCTC — you used our services last year. If you haven't filed yet this year, you might still be eligible to claim tax credits like the stimulus! Sign up by November 15 at GetCTC.org/Nov15-File	22,845 recipients Hello! This is GetCTC — you used our services last year. If you haven't filed yet this year, you might be eligible to claim cash benefits like the stimulus! Sign up by November 15 at GetCTC.org/Nov15File
	No deadline	22,845 recipients Hello! This is GetCTC — you used our services last year. If you haven't filed yet this year, you might still be eligible to claim tax credits like the stimulus! Sign up at GetCTC.org/Nov-File	22,845 recipients Hello! This is GetCTC — you used our services last year. If you haven't filed yet this year, you might be eligible to claim cash benefits like the stimulus! Sign up at GetCTC.org/NovFile
Cohort 2	Deadline	36,274 recipients Hello! This is GetCTC — you signed up for updates about our services. If you haven't filed yet this year, you might still be eligible to claim tax credits like the stimulus! Sign up by November 15 at GetCTC.org/Nov15_File	36,274 recipients Hello! This is GetCTC — you signed up for updates about our services. If you haven't filed yet this year, you might be eligible to claim cash benefits like the stimulus! Sign up by November 15 at GetCTC.org/fileNov15
	No deadline	36,274 recipients Hello! This is GetCTC — you signed up for updates about our services. If you haven't filed yet this year, you might still be eligible to claim tax credits like the stimulus! Sign up at GetCTC.org/Nov_File	36,274 recipients Hello! This is GetCTC — you signed up for updates about our services. If you haven't filed yet this year, you might be eligible to claim cash benefits like the stimulus! Sign up at GetCTC.org/fileNov

Messages were sent on November 9.

Messages to Cohort 1 were sent through The Hub, GetCTC's client interaction system, in the same vein as any other messages sent to clients, including updates on return status and customer service interactions with clients experiencing issues. Messages sent as text messages would appear from Code for America's shortcode, in the same thread with all other messages from the product. This would be a trusted phone number for the recipients.

Messages to Cohort 2 were also sent as text messages from Code for America's shortcode, akin to messages from The Hub. Since these individuals did not have ongoing client relationships with Code for America, however, the recipients may not have recognized the shortcode as belonging to Code for America, and the messages would not have arrived on the same thread with other messages from our services—other than the earlier outreach nudge in May.

As in other studies on Code for America products this year, outcomes for Cohort 1 can be calculated in two principal ways: (1) by matching the Social Security Number of the individual receiving the outreach to the Social Security Number on GetCTC returns, or (2) by looking at returns filed using specific source URLs (e.g., getctc.org/claimnow and getctc.org/file-now), which were assigned to the groups in outgoing messages and can be tracked through to completion. The former is a more precise count of how many clients actually submitted returns. The latter, on the other hand, allows us to look farther up the funnel at home page visits (which occur prior to entering SSN). The latter may also abstract away from some amount of noise, since clients in this population were subject to various other messages during this period; the source URLs capture something like the effect on filing *by this outreach*.

For Cohort 2, however, SSN matching is not an option because the SSN of the recipient is not known. Cohort 2 does allow for outcome measuring by matching the phone number the client entered against the phone number on a GetCTC return, which is, like SSN, probably a more precise measure of the number of returns actually submitted.

In the results below, we show pooled results by phone and source, as well as results by SSN for Cohort 1 only.

3. Results

3.1 Tax language

Table 1 shows the principal results for the tax language experiment. The different specifications all tell a relatively similar story: Using non-tax language significantly increases home page visits (Column 1), and may increase starts (Column 5) and submissions (Columns 6 and 9), but it does not have a detectable impact on the number of accepted returns generated. Appendix Table A1 disaggregates Panel A of this table by cohort, and Appendix Table A2 disaggregates Panel B of this table by cohort; neither find meaningful differences between the cohorts.

Note that the absolute counts vary very significantly across methods, with source picking up far fewer returns than phone or SSN. The discrepancy is due to the large number of outreach campaigns that

were occurring at the same time, during the last week of GetCTC operation. Many of the in-sample clients who filed did so using a link they received from another outreach campaign.

Table 1 - Headline results for tax versus non-tax language

	Pooled cohorts							Cohort 1 only		
	A. By Source				B. By Phone			C. By SSN		
	Home view	Start	Subm	Acc	Start	Subm	Acc	Start	Subm	Acc
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Tax language	4.34% (5127)	0.58% (685)	0.231% (273)	0.082% (97)	2.00% (2359)	0.929% (1099)	0.303% (358)	2.17% (991)	1.00% (456)	0.370% (169)
Non-tax language	5.12% (6059)	0.60% (708)	0.223% (264)	0.082% (97)	2.11% (2493)	1.005% (1188)	0.311% (368)	2.21% (1012)	1.09% (497)	0.328% (150)
P(T = C)	<.0001	0.53	0.68	1.0	0.059	0.059	0.73	0.63	0.181	0.28

Note: 'Start' is defined as entering an SSN.

3.2 Deadline language

Table 2 shows the headline results for the deadline language experiment. The story is relatively similar to the story with tax language. Deadline language significantly increases home page views (Column 1), is very likely to significantly increase return starts (Columns 2, 5, 8), and may significantly increase return submissions (Column 3, 9). But, aside from the weakly suggestive result in Column 4 (p=.17), there is again no evidence of an impact on the number of accepted returns. The lack of an effect on accepted returns is perhaps more striking here than in the tax language experiment, given the clearer evidence of an effect through the rest of the funnel. Appendix Table A3 disaggregates Panel A of this table by cohort and Appendix Table A4 disaggregates Panel B of this table by cohort; neither uncover meaningful discrepancies across cohorts.

Table 2 - Headline results for deadline versus non-deadline language

	Pooled cohorts							Cohort 1 only		
	A. By Source				B. By Phone			C. By SSN		
	Home view	Start	Subm	Acc	Start	Subm	Acc	Start	Subm	Acc
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Deadline language	5.10% (6026)	0.65% (773)	0.244% (288)	0.090% (107)	2.15% (2540)	0.979% (1158)	0.304% (360)	2.26% (1032)	1.09% (499)	0.346% (158)

No dead. language	4.36% (5160)	0.52% (620)	0.211% (249)	0.074% (87)	1.96% (2312)	0.955% (1129)	0.310% (366)	2.13% (971)	0.99% (454)	0.352% (161)
P(T = C)	<.0001	<.0001	0.092	0.17	0.001	0.55	0.79	0.167	0.145	0.88

Note: 'Start' is defined as entering an SSN.

4. Discussion

The results confirm earlier findings and findings from behavioral research that non-tax language and deadline language increase site visits and action rates. But neither set of language has indisputable impact on submissions, and neither has any appreciable impact on acceptances.

There are, broadly, two ways to interpret the results. One interpretation is that the non-tax and deadline language clearly performs better, but the effect grows too small at the end of the funnel to be detectable in this study. (We were powered to detect small effects in at least some specifications; at a baseline acceptance rate of 0.3%, we were powered to detect a 0.075 percentage point effect. The low baseline acceptance rate means that the tiny absolute effect is non-negligible in relative terms.) The other is that the higher visit rate is real but effectively meaningless. The non-tax and deadline language drives additional people to the site, but at the margin, none of them are actually going to file a tax return. In fact, in the tax language case, perhaps many of the message recipients know that they already filed a return and accessed their tax benefits, and they are induced to click simply because the message confusingly appears to advertise new benefits that they did not know about. In this sense the message would be just attention-catching spam and, if anything, could be damaging, by creating an environment of misleading salesmanship.

Regardless of the interpretation, it is hard to escape the conclusion that such messaging choices do not have a large effect on acceptances one way or the other.

At a broader level, the results offer a word of caution against messaging experiments that do not measure outcomes past the top of the funnel. It may seem reasonable enough to simply measure click-through rates, assuming they function as a proxy for true outcomes of interest (e.g., filing a return). But there is no such guarantee. In these cases, the better-performing language drives meaningfully more site visits—about 18% more, or an additional 900 visitors. Were this study to end at the level of the site visit, one would conclude that non-tax language and deadline language are clearly preferable. But the results vanish entirely at the end of the funnel. Studies that end too far up the funnel may well find plenty of eye-catching results that end up being meaningless.

Appendix

Table A1 - Results for tax versus non-tax language, source match, by cohort (Table 1, Panel A, disaggregated)

	By Source							
	A. Cohort 1				B. Cohort 2			
	Homep	Start	Subm	Acc	Homep	Start	Subm	Acc
Tax language	3.46% (1581)	0.71% (324)	0.31% (141)	.118% (54)	4.89% (3546)	0.50% (361)	0.16% (117)	.061% (44)
Non-tax language	4.08% (1866)	0.74% (336)	0.32% (147)	.116% (53)	5.78% (4193)	0.51% (372)	0.18% (132)	.059% (43)
P(T = C)	<0.0001	0.64	0.73	0.93	<0.0001	0.69	0.33	0.88

Note: 'Start' is defined as entering an SSN.

Table A2 - Results for tax versus non-tax language, phone match (Table 1, Panel B, disaggregated)

	By Phone					
	Cohort 1			Cohort 2		
	Start	Subm	Acc	Start	Subm	Acc
Tax language	1.52% (695)	0.72% (330)	.306% (140)	2.29% (1664)	1.06% (769)	.300% (218)
Non-tax language	1.60% (729)	0.79% (359)	.271% (124)	2.43% (1764)	1.14% (829)	.336% (244)
P(T = C)	0.33	0.22	0.32	0.079	0.13	0.22

Note: 'Start' is defined as entering an SSN.

Table A3 - Results for deadline versus non-deadline language, source match, by cohort (Table 2, Panel A, disaggregated)

	By Source							
	Cohort 1				Cohort 2			
	Homep	Start	Subm	Acc	Homep	Start	Subm	Acc
Deadline language	4.12% (1883)	0.78% (358)	0.35% (159)	.123% (56)	5.71% (4143)	0.57% (415)	0.18% (129)	.070% (51)

No deadline lang	3.42% (1564)	0.66% (302)	0.28% (129)	.112% (51)	4.96% (3596)	0.44% (318)	0.17% (120)	.050% (36)
P(dead = no dead)	<0.0001	0.028	0.075	0.63	<0.0001	0.0003	0.55	0.12

Note: 'Start' is defined as entering an SSN.

Table A4 - Results for deadline versus non-deadline language, phone match (Table 2, Panel B, disaggregated)

	By Phone					
	Cohort 1			Cohort 2		
	Start	Subm	Acc	Start	Subm	Acc
Deadline language	1.60% (729)	0.79% (363)	.289% (132)	2.50% (1811)	1.10% (795)	.314% (228)
No deadline lang	1.52% (695)	0.71% (326)	.289% (132)	2.23% (1617)	1.11% (803)	.323% (234)
P(tax = non-tax)	0.33	0.16	1.0	0.0007	0.84	0.76

Note: 'Start' is defined as entering an SSN.