Can you spot a scam?
Measuring and improving scam identification ability

Elif Kubilay[1] ⓡ Eva Raiber[2] ⓡ Lisa Spantig[3,1] ⓡ Jana Cahlkov[4] ⓡ Lucy Kaaria[5]

[1]University of Essex,[2]Aix-Marseille Université, CNRS, AMSE

[3]RWTH Aachen University,[4]University of Bonn,[5]University of Nairobi and HOPAWI

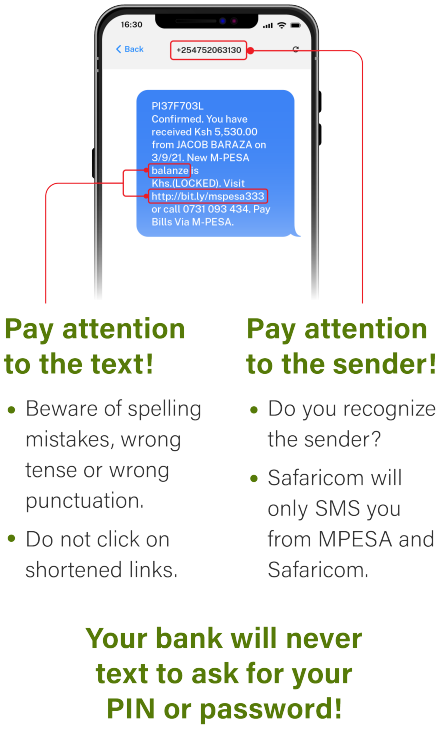
Abstract

The expansion of digital financial services leads to severe consumer protection issues such as fraud and scams. As these potentially decrease trust in digital services, especially in developing countries, avoiding victimization has become an important policy objective. In an online experiment, we first investigate how well individuals in Kenya identify phone scams using a novel measure of scam identification ability. We then test the effectiveness of scam education, a commonly used approach by organizations for fraud prevention. We find that common tips on how to spot scams do not significantly improve individuals' scam identification ability, i.e., the distinction between scams and genuine messages. This null effect is driven by an increase in correctly identified scams and a decrease in correctly identified genuine messages, indicating overcaution. Additionally, we find suggestive evidence that genuine messages with scam-like features are misclassified more often, highlighting the importance of a careful design of official communication.

JEL Codes: D14, D18, G53, O12

Keywords: Consumer protection, consumer fraud, digital financial services, scam susceptibility, scam education, Kenya

Figure 1: Tips treatment



Notes: Tips treatment was designed based on commonly communicated tips in Kenya. The graphic was "animated," such that the pieces of information would be shown step-by-step. Participants clicked through this animation at their own speed, i.e., they hit the "continue" button five times before they see the overall graphic.

Table 1: Correlates of Scam Identification Ability and Confidence

	SIA			Confidence in SIA		
	(1)	(2)	(3)	(1)	(2)	(3)
Demographics:						
Female	-0.03***	-0.03***	-0.03**	-0.11***	-0.10***	-0.10**
	(0.01)	(0.01)	(0.01)	(0.04)	(0.04)	(0.04)
Age in Years	0.00	0.00	0.00	0.01**	0.01**	0.01**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Post-Seconday Education	0.03*	0.02	0.02	0.11**	0.10*	0.09*
	(0.01)	(0.01)	(0.01)	(0.05)	(0.05)	(0.05)
Low Income	0.02	0.02	0.02	0.04	0.03	0.03
	(0.01)	(0.01)	(0.01)	(0.05)	(0.05)	(0.05)
Formal Employment	-0.00	-0.01	-0.00	0.06	0.04	0.05
	(0.01)	(0.01)	(0.01)	(0.04)	(0.04)	(0.04)
DFS Use:						
Low Trust in DFS		0.01	0.01		-0.11**	-0.11**
		(0.01)	(0.01)		(0.05)	(0.05)
High use of different DFS		0.03**	0.03**		0.05	0.05
		(0.01)	(0.01)		(0.04)	(0.04)
Scam Experience:						
Contacted less than 1 week ago			-0.01			0.00
			(0.02)			(0.06)
Victim of a Scammer			-0.00			-0.02
			(0.01)			(0.04)
N	997	997	991	997	997	991
R-Squared	0.05	0.05	0.05	0.03	0.04	0.03

Notes: Dependent variables are the share of correctly identified messages (SIA) in block 1 and average confidence ratings in block 1. *Female, Post-Secondary Education, Formal Employment, Low Trust in DFS, Contacted less than 1 week ago*, and *victim of a scammer* are binary indicators, *Low Income* and *High use of different DFS* are binary indicators for median splits. All variables rely on self-reports. All specifications control for the order of the two blocks and failing the attention check. The displayed coefficients are from OLS regressions. Robust standard errors are in parenthesis. Asterisks indicate that the estimate is statistically significant at the 1% ***, 5% **, and 10% * levels.

Table 2: Treatment Effects

Panel 1: Main Outcomes

	Correctly Identified Messages			Confidence		
	SIA	Scams	Non-scams	SIA	Scams	Non-scams
Tips (unincentivized)	0.02	0.08***	-0.09***	0.12***	0.16***	0.06
	(0.02)	(0.02)	(0.03)	(0.04)	(0.05)	(0.05)
Tips (incentivized)	0.03*	0.08***	-0.07**	0.08*	0.08	0.07
	(0.02)	(0.02)	(0.03)	(0.04)	(0.05)	(0.06)
Control Mean	0.70	0.69	0.71	4.20	4.13	4.33
p-value ($Tips^U = Tips^I$)	0.69	0.82	0.60	0.37	0.14	0.85
N	991	991	991	991	991	991
R-Squared	0.04	0.11	0.16	0.47	0.40	0.27

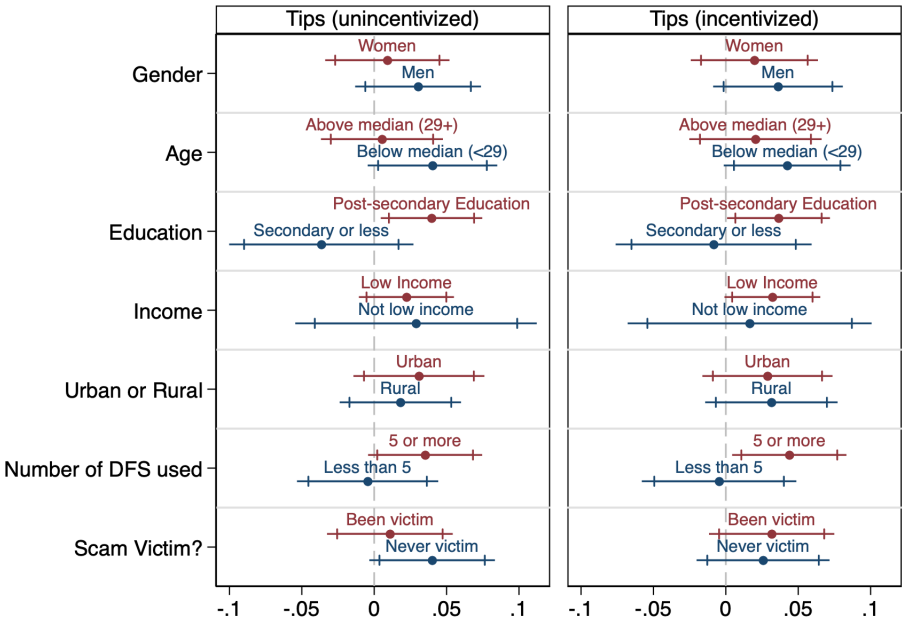
Panel 2: Secondary Outcomes

	Low Trust in DFS	Response Time SIA	All Scams Identified	All Non-scam Identified
Tips (unincentivized)	-0.01	0.11	0.10**	-0.11**
	(0.04)	(0.08)	(0.04)	(0.04)
Tips (incentivized)	-0.02	0.21**	0.11***	-0.08*
	(0.04)	(0.09)	(0.04)	(0.04)
Control Mean	0.32	2.21	0.30	0.52
p-value ($Tips^U = Tips^I$)	0.92	0.27	0.82	0.48
N	991	991	991	991
R-Squared	0.04	0.34	0.07	0.11

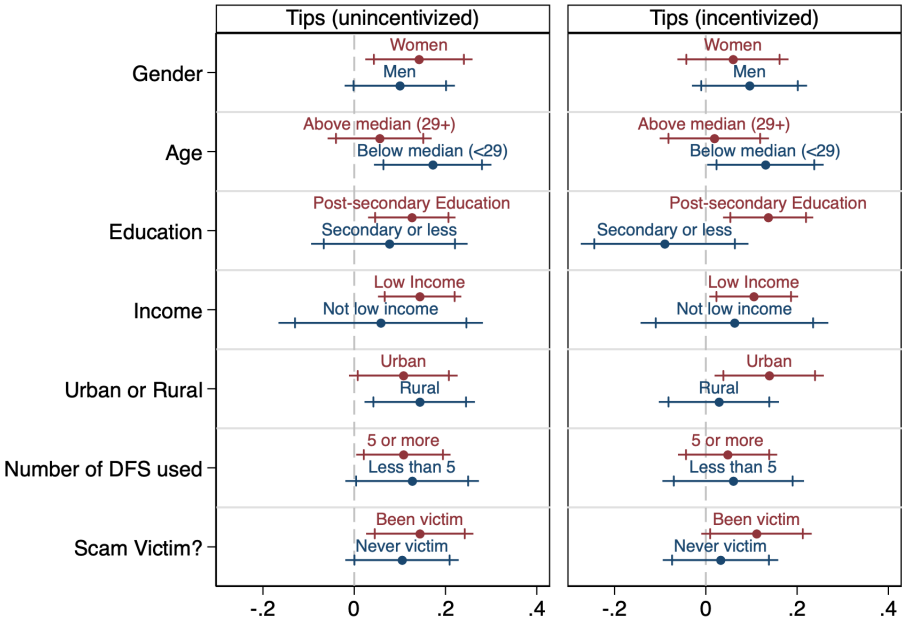
Notes: In Panel 1, the dependent variables are the share of correctly identified messages(SIA) in block 2, the share of correctly identified scams in block 2, the share of correctly identified non-scams in block 2, and the average confidence ratings in block 2 for all messages (confidence in SIA), for the scam messages, and for the non-scam messages. In Panel 2, the dependent variables are a binary indicator for low trust in DFS, the time spent on SIA in block 2, a binary indicator for classifying all scams correctly in block 2, and a binary indicator for classifying all non-scams correctly in block 2. All specifications include an indicator for the incentives treatment, the value of the outcome variable in block 1 (except for trust, which was only measured after block 2), and the full set of controls, i.e., variables displayed in Table 1 (female, age, post-secondary education, low income, formal employment, low trust in DFS (except for the effect on trust), above median use of different DFS, contacted less than one week ago, victim of a scammer), as well as indicators for the order of the two blocks and failing the attention check. $Tips^U$ and $Tips^I$ refer to Tips (unincentivized) and Tips (incentivized), respectively. The displayed coefficients are from OLS regressions. Robust standard errors are in parenthesis. Asterisks indicate that the estimate is statistically significant at the 1% ***, 5% **, and 10% * levels.

Figure 2: Treatment Effect Heterogeneity

(a) Scam Identification Ability (SIA)

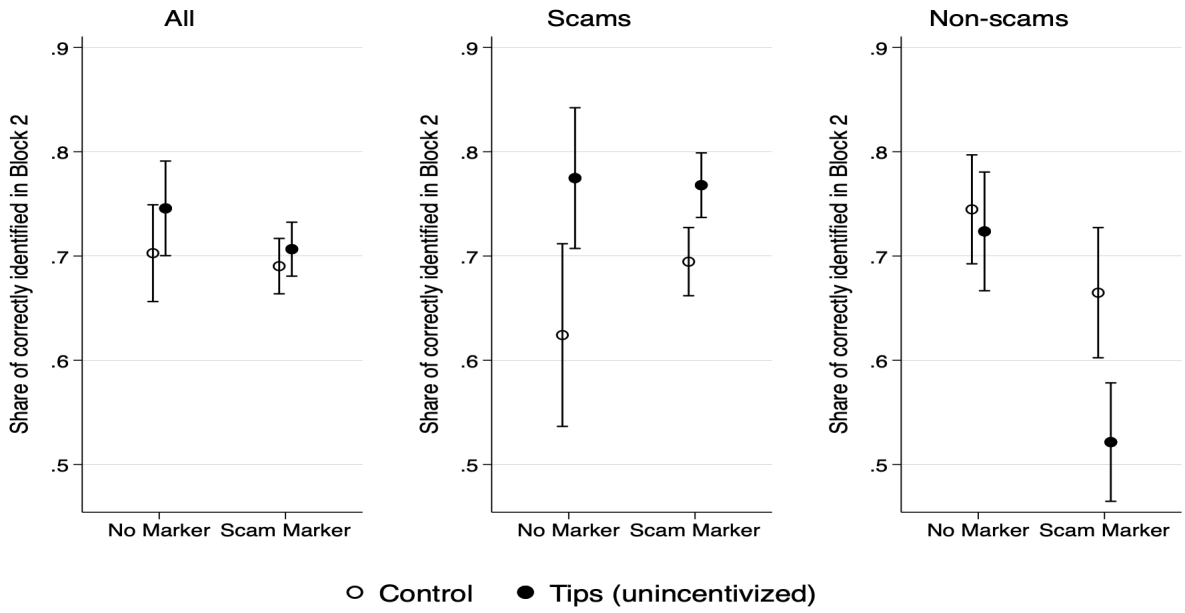


(b) Confidence



Notes: Figures plot the OLS coefficients and the 90% and 95% confidence intervals from the estimating regressions in Panel 1, Table 2 (Column 1 for SIA and Column 4 for Confidence) separately for the different subcategories.

Figure 3: Vignette-level effects by whether the message contains a scam marker



Notes: Figures plot the average marginal effects of triple-differences estimation with 95% confidence intervals based on standard errors clustered at the respondent level (see also Appendix D). Scam Marker is an indicator for whether the message contains at least one of the scam markers the tips warn about. The left panel contains all vignettes, the center panel focuses on scams, and the right panel on non scams. For ease of exposition, only the control and the Tips (unincentivized) treatment are displayed. The empirical specification contains the full set of interactions and demographic controls.

Online Appendix

A Additional Tables

Table A1: Overview of vignettes

	Content	Intention	Sender
Block A	M-PESA transfer receipt	Genuine	Displayed
	Offer to use the new M-PESA app and get cash back	Genuine	Not displayed
	Random message to encourage contact	Fraudulent	Displayed
	Investment opportunity	Fraudulent	Displayed
	Suspended bank account	Fraudulent	Not displayed
	Notification as emergency contact	Fraudulent	Not displayed
Block B	M-PESA reversal request	Genuine	Displayed
	Notification of new registered SIM	Genuine	Displayed
	Job offer	Fraudulent	Displayed
	Lottery win	Fraudulent	Displayed
	Covid-19 relief fund	Fraudulent	Not displayed
	Notification as loan grantor	Fraudulent	Displayed

Notes: Vignettes kept the original wording of the screenshots and were visually harmonized, i.e., all vignettes were displayed on the same phone with the same signal strength, battery level, etc. (see also Figure A1). The order of the blocks was randomized at the individual level, as was the order of vignettes within a block. As blocks might be different for various reasons, we always control for the order of the blocks in the analysis.

Table A2: Balance

	(1) Control		(2) Tips (unincentivized)		(3) Incentives		(4) Tips (incentivized)		F-test for balance across all groups		(1)-(2)		(1)-(3)		(1)-(4)		(2)-(3)		(2)-(4)		(3)-(4)	
											Pairwise t-test											
Variable	N	Mean/(SE)	N	Mean/(SE)	N	Mean/(SE)	N	Mean/(SE)	N	F-stat/P-value	N	P-value	N	P-value	N	P-value	N	P-value	N	P-value	N	P-value
Female (0/1)	256	0.52	259	0.50	246	0.50	239	0.48	1000	0.20	515	0.76	502	0.80	495	0.44	505	0.96	498	0.64	485	0.62
		(0.03)		(0.03)		(0.03)		(0.03)		0.90												
Age	256	32.23	259	32.55	246	32.07	239	32.26	1000	0.11	515	0.72	502	0.85	495	0.97	505	0.58	498	0.76	485	0.83
		(0.61)		(0.62)		(0.59)		(0.67)		0.96												
Urban (0/1)	256	0.47	259	0.47	246	0.52	238	0.53	999	1.16	515	0.97	502	0.21	494	0.18	505	0.20	497	0.17	484	0.91
		(0.03)		(0.03)		(0.03)		(0.03)		0.32												
Post secondary education (0/1)	256	0.78	259	0.74	246	0.71	239	0.70	1000	1.62	515	0.34	502	0.09*	495	0.05**	505	0.45	498	0.29	485	0.76
		(0.03)		(0.03)		(0.03)		(0.03)		0.18												
Low income (0/1)	256	0.80	259	0.81	246	0.74	239	0.77	1000	1.75	515	0.69	502	0.09*	495	0.40	505	0.03**	498	0.22	485	0.38
		(0.03)		(0.02)		(0.03)		(0.03)		0.15												
Formal employment (0/1)	256	0.36	258	0.38	245	0.37	238	0.35	997	0.14	514	0.77	501	0.92	494	0.74	503	0.84	496	0.53	483	0.67
		(0.03)		(0.03)		(0.03)		(0.03)		0.94												
Internet on phone (0/1)	256	0.99	258	0.99	246	1.00	239	0.99	999	0.36	514	0.66	502	0.58	495	0.95	504	0.33	497	0.71	485	0.55
		(0.01)		(0.01)		(0.00)		(0.01)		0.78												
Social media on phone (0/1)	256	0.99	258	0.99	246	0.99	239	1.00	999	0.49	514	0.66	502	0.62	495	0.60	504	0.95	497	0.35	485	0.33
		(0.01)		(0.01)		(0.01)		(0.00)		0.69												
Recent DFS use (0/1)	251	0.95	253	0.94	240	0.95	236	0.97	980	0.78	504	0.70	491	0.92	487	0.30	493	0.63	489	0.16	476	0.36
		(0.01)		(0.01)		(0.01)		(0.01)		0.50												
Number of DFS used	256	4.85	259	4.79	246	4.68	239	4.79	1000	0.19	515	0.80	502	0.47	495	0.81	505	0.63	498	0.99	485	0.62
		(0.16)		(0.16)		(0.16)		(0.16)		0.90												
Above median use of DFS (0/1)	256	0.62	259	0.60	246	0.61	239	0.63	1000	0.24	515	0.60	502	0.72	495	0.81	505	0.87	498	0.45	485	0.55
		(0.03)		(0.03)		(0.03)		(0.03)		0.87												

Notes: Asterisks indicate that the difference is statistically significant at the 1% ***, 5% **, and 10% * levels.

Table A3: Sample and Kenyan Population

	Online Survey	**Kenya National Bureau of Statistics (2019)**
	Fraction	Fraction of total population (adults)
Gender		
Female	50.1%	50.5%
Male	49.9%	49.5%
Age		
18-24	32.1%	13.4% (24.8%)
25-34	27.0%	15.6% (29.0%)
35+	40.9%	24.9% (46.3%)
County		
Baringo	1.4%	1.4%
Bomet	1.9%	1.8%
Bungoma	3.7%	3.5%
Busia	2.0%	1.9%
Elgeyo-Marakwet	1.0%	1.0%
Embu	1.3%	1.3%
Garissa	1.6%	1.8%
Homa Bay	2.5%	2.4%
Isiolo	0.4%	0.6%
Kajiado	1.7%	2.4%
Kakamega	4.4%	3.9%
Kericho	1.9%	1.9%
Kiambu	4.4%	5.1%
Kilifi	3.0%	3.1%
Kirinyaga	1.4%	1.3%
Kisii	2.9%	2.7%
Kisumu	2.5%	2.4%
Kitui	2.6%	2.4%
Kwale	1.7%	1.8%
Laikipia	1.0%	1.1%
Lamu	0.3%	0.3%
Machakos	2.8%	3.0%
Makueni	2.4%	2.1%
Mandera	2.5%	1.8%
Marsabit	0.8%	1.0%
Meru	3.4%	3.2%
Migori	2.4%	2.3%
Mombasa	2.4%	2.5%
Murang'a	2.4%	2.2%
Nairobi	8.3%	9.2%
Nakuru	4.2%	4.5%
Nandi	2.0%	1.9%
Narok	2.2%	2.4%
Nyamira	1.5%	1.3%
Nyandarua	1.5%	1.3%
Nyeri	1.8%	1.6%
Samburu	0.7%	0.7%
Siaya	2.1%	2.1%
Taita-Taveta	0.7%	0.7%
Tana River	0.6%	0.7%
Tharaka-Nithi	0.9%	0.8%
Trans Nzoia	2.1%	2.1%
Turkana	2.1%	1.9%
Uasin Gishu	2.3%	2.4%
Vihiga	1.4%	1.2%
Wajir	1.6%	1.6%
West Pokot	1.3%	1.3%
Location		
Urban	49.6%	31.2%
Rural	50.3%	68.8%

Table A4: Sample Characteristics

	N	Mean	SD	Min.	Max.
Demographics					
Female	1000	0.50	0.50	0	1
Age	1000	32.28	9.84	18	67
Urban	999	0.50	0.50	0	1
Post secondary education	1000	0.73	0.44	0	1
Low income	1000	0.78	0.41	0	1
Formal employment	997	0.36	0.48	0	1
Internet on phone	999	0.99	0.09	0	1
Social media on phone	999	0.99	0.09	0	1
Financial transactions w/ phone in the past 90 days	980	0.96	0.21	0	1
DFS Use					
Number of DFS used	1000	4.78	2.52	0	9
Scam Experience					
Have you ever been contacted by a scammer?	999	0.96	0.18	0	1
Ever been a victim of a scammer?	995	0.54	0.50	0	1
Anyone you know ever been a victim of a scammer?	1000	0.85	0.35	0	1
Scam Identification Ability (Block 1)					
Share of correctly identified messages (SIA)	1000	0.71	0.18	0	1
Share of correctly identified scams	1000	0.74	0.24	0	1
Share of correctly identified non-scams	1000	0.66	0.35	0	1
Average confidence in SIA	1000	4.23	0.63	1	5

Table A5: The Effects of Incentives

Panel 1: Outcomes in Block 1

	SIA	Scams Identified	Non-scams Identified	Confidence
Incentives	-0.01	-0.01	0.00	0.06
	(0.02)	(0.02)	(0.03)	(0.06)
Control Mean	0.71	0.75	0.63	4.23
N	991	991	991	991
R-Squared	0.05	0.04	0.09	0.04

Panel 2: Outcomes in Block 2

	SIA	Scams Identified	Non-scams Identified	Confidence
Incentives	0.02	0.02	0.02	0.11**
	(0.02)	(0.02)	(0.03)	(0.05)
Control Mean	0.70	0.69	0.71	4.20
N	991	991	991	991
R-Squared	0.03	0.07	0.12	0.05

Notes: In Panels 1 and 2, dependent variables are the share of correctly identified messages (SIA), the share of correctly identified scams, the share of correctly identified non-scams, and the average confidence ratings in block 1 and 2, respectively. All specifications include indicators for the tips treatments, and the full set of controls, i.e., variables displayed in Table 1 (female, age, post-secondary education, low income, formal employment, low trust in DFS, above median use of different DFS, contacted less than one week ago, victim of a scammer), as well as indicators for the order of the two blocks and failing the attention check. The displayed coefficients are from OLS regressions. Robust standard errors are in parenthesis. Asterisks indicate that the estimate is statistically significant at the 1% ***, 5% **, and 10% * levels.

Table A6: Confidence weighted by SIA

	All messages	Scams	Non-scams
Tips (unincentivized)	0.05**	0.15***	-0.16***
	(0.03)	(0.04)	(0.05)
Tips (incentivized)	0.06**	0.15***	-0.12**
	(0.03)	(0.04)	(0.05)
Control Mean	0.36	0.34	0.38
p-value ($Tips^U = Tips^I$)	0.88	0.95	0.53
N	991	991	991
R-Squared	0.05	0.13	0.17

Notes: The dependent variables are weighted confidence in block 2 for all messages, scams, and non-scams, respectively. Weighted confidence ranges from -1 to 1, where 1 means fully confident and perfect SIA, whereas -1 means fully confident and no correctly classified vignette. All specifications include an indicator for the incentives treatment, the value of the outcome variable in block 1, and the full set of controls, i.e., variables displayed in Table 1 (female, age, post-secondary education, low income, formal employment, low trust in DFS, above median use of different DFS, contacted less than one week ago, victim of a scammer), as well as indicators for the order of the two blocks and failing the attention check. $Tips^U$ and $Tips^I$ refer to Tips (unincentivized) and Tips (incentivized), respectively. The displayed coefficients are from OLS regressions. Robust standard errors are in parenthesis. Asterisks indicate that the estimate is statistically significant at the 1% ***, 5% **, and 10% * levels.

Additional Figures

Figure A1: Example Vignette

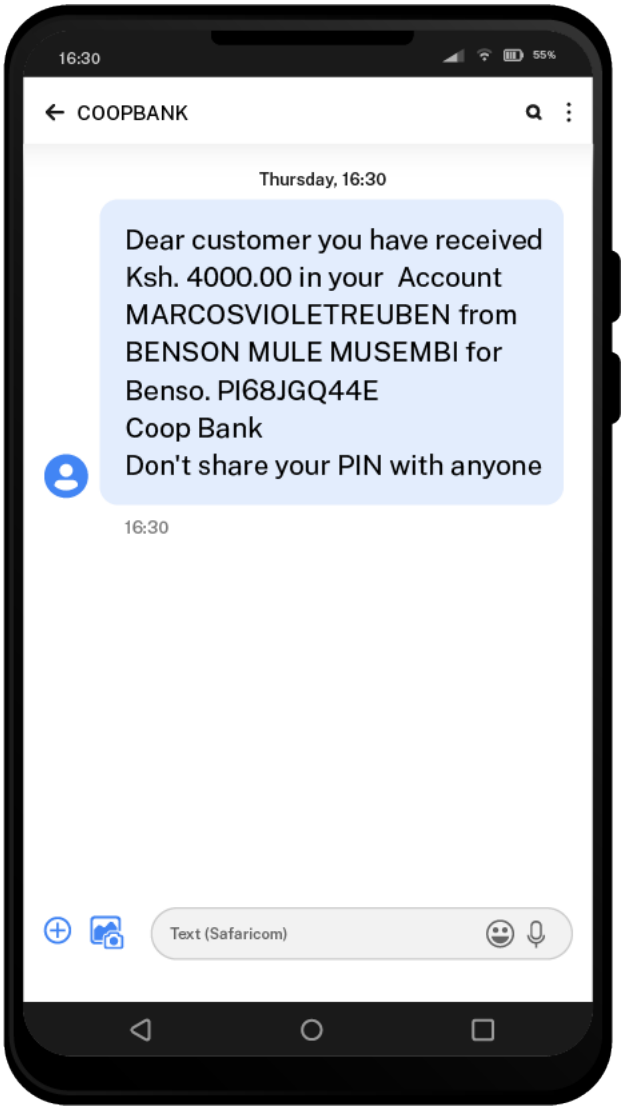
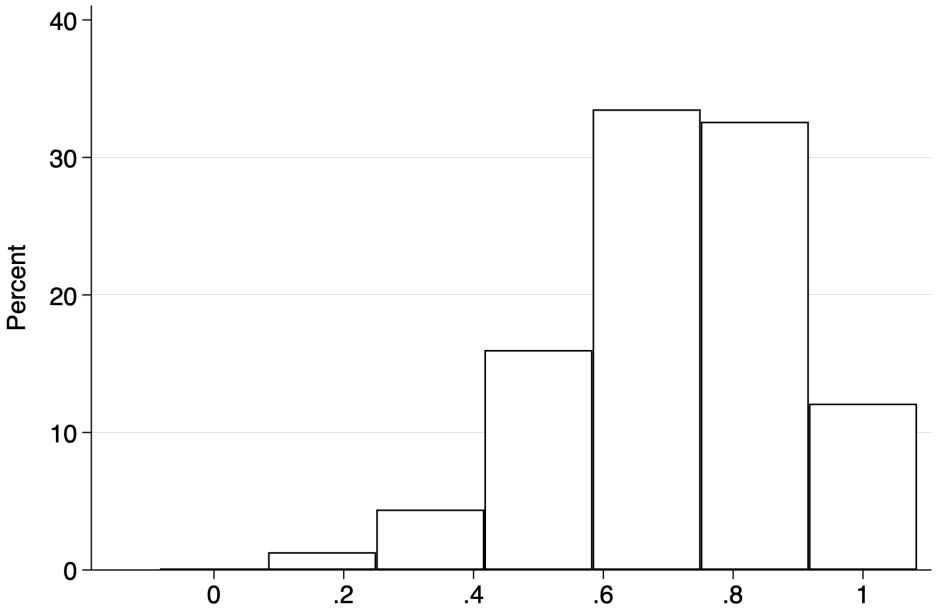
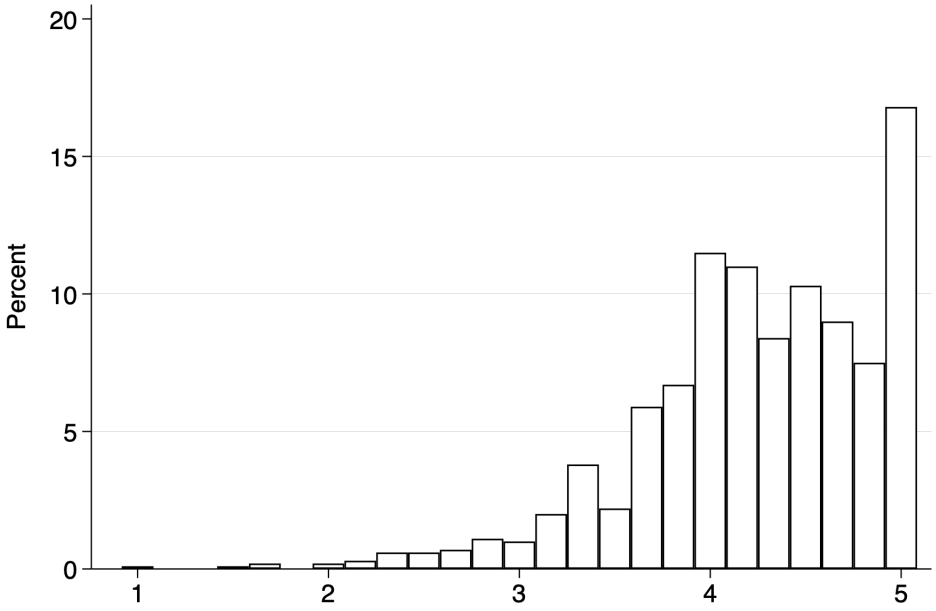


Figure A2: SIA and Confidence in Block 1

(a) Distribution of scam identification ability (SIA)



(b) Distribution of Confidence in SIA



B Robustness

Table B1: Attention check and main treatment effects

	Passed Attention Check (SIA)			Passed Attention Check (Confidence)			All
	All	Scams	Non-scams	All	Scams	Non-scams	Failed Attention
Tips (unincentivized)	0.04*	0.10***	-0.09***	0.15***	0.20***	0.03	-0.01
	(0.02)	(0.03)	(0.03)	(0.05)	(0.06)	(0.07)	(0.04)
Tips (incentivized)	0.03	0.09***	-0.10***	0.05	0.04	0.05	-0.09**
	(0.02)	(0.03)	(0.03)	(0.05)	(0.06)	(0.07)	(0.04)
Control Mean	0.69	0.67	0.73	4.18	4.10	4.33	0.30
p-value ($Tips^U = Tips^I$)	0.59	0.73	0.86	0.07	0.01	0.80	0.04
N	727	727	727	727	727	727	991
R-Squared	0.06	0.12	0.17	0.46	0.39	0.27	0.05

Notes: Columns 1-6 include only those who passed the attention check. Column 7 includes all participants. Dependent variables are the share of correctly identified messages (SIA) in block 2, the share of correctly identified scams in block 2, the share of correctly identified non-scams in block 2, the average confidence ratings in block 2 for all messages (confidence in SIA), for scam messages, and for non-scam messages and an indicator for failing the attention check. All specifications include an indicator for the incentives treatment, the value of the outcome variable in block 1 (except for the attention check which was only administered once), and the full set of controls, i.e., variables displayed in Table 1 (female, age, post-secondary education, low income, formal employment, low trust in DFS, above median use of different DFS, contacted less than one week ago, victim of a scammer), as well as indicators for the order of the two blocks and failing the attention check (except for the last specification where the attention check is the dependent variable). $Tips^U$ and $Tips^I$ refer to Tips (unincentivized) and Tips (incentivized), respectively. The displayed coefficients are from OLS regressions. Robust standard errors are in parenthesis. Asterisks indicate that the estimate is statistically significant at the 1% ***, 5% **, and 10% * levels.

Table B2: Treatment Effects: Additional Robustness Checks

Panel 1: Excluding the Baseline SIA and Confidence

	Correctly Identified Messages			Confidence		
	SIA	Scams	Non-scams	SIA	Scams	Non-scams
Tips (unincentivized)	0.02	0.08***	-0.08***	0.13**	0.16***	0.07
	(0.02)	(0.02)	(0.03)	(0.06)	(0.06)	(0.06)
Tips (incentivized)	0.03*	0.07***	-0.06**	0.05	0.06	0.04
	(0.02)	(0.02)	(0.03)	(0.06)	(0.07)	(0.07)
Control Mean	0.70	0.69	0.71	4.20	4.13	4.33
p-value ($Tips^U = Tips^I$)	0.79	0.86	0.51	0.21	0.13	0.70
N	991	991	991	991	991	991
R-Squared	0.03	0.07	0.12	0.05	0.05	0.04

Panel 2: Incentivized Sample

	Correctly Identified Messages			Confidence		
	SIA	Scams	Non-scams	SIA	Scams	Non-scams
Tips (incentivized)	0.01	0.06**	-0.09***	0.02	0.01	0.01
	(0.02)	(0.02)	(0.03)	(0.04)	(0.05)	(0.06)
Incentives (no-tips) Mean	0.71	0.71	0.73	4.30	4.24	4.40
N	479	479	479	479	479	479
R-Squared	0.06	0.12	0.14	0.51	0.46	0.27

Panel 3: Unincentivized Sample

	Correctly Identified Messages			Confidence		
	SIA	Scams	Non-scams	SIA	Scams	Non-scams
Tips (unincentivized)	0.02	0.08***	-0.09***	0.12***	0.16***	0.05
	(0.02)	(0.02)	(0.03)	(0.04)	(0.05)	(0.06)
Control Mean	0.70	0.69	0.71	4.20	4.13	4.33
N	512	512	512	512	512	512
R-Squared	0.06	0.11	0.19	0.44	0.37	0.28

Notes: The dependent variables are the scam identification ability share of correctly identified messages (SIA) in block 2, the share of correctly identified scams in block 2, the share of correctly identified non-scams in block 2, and the average confidence ratings in block 2 for all messages (confidence in SIA), for the scam messages, and for the non-scam messages. In Panel 1, all specifications include an indicator for the incentives treatment. In Panels 2 and 3, the sample is restricted to incentivized and unincentivized respondents, respectively. All specifications include the full set of controls, i.e., variables displayed in Table 1 (female, age, post-secondary education, low income, formal employment, low trust in DFS, above median use of different DFS, contacted less than one week ago, victim of a scammer), as well as indicators for the order of the two blocks and failing the attention check. $Tips^U$ and $Tips^I$ refer to Tips (unincentivized) and Tips (incentivized), respectively. Displayed coefficients are from OLS regressions. Robust standard errors are in parenthesis. Asterisks indicate that the estimate is statistically significant at the 1% ***, 5% **, and 10% * levels.

Table B3: Treatment Effects: Varying Control Variables

Panel 1: Scam Identification Ability

	(1)	(2)	(3)	(4)
Tips (unincentivized)	0.02	0.02	0.02	0.02
	(0.02)	(0.02)	(0.02)	(0.02)
Tips (incentivized)	0.02	0.03*	0.03*	0.03*
	(0.02)	(0.02)	(0.02)	(0.02)
Demographics		✓	✓	✓
Design Controls			✓	✓
Scam Experience				✓
Control Mean	0.70	0.70	0.70	0.70
p-value ($Tips^U = Tips^I$)	0.81	0.59	0.70	0.69
N	1000	997	997	991
R-Squared	0.01	0.03	0.04	0.04

Panel 2: Confidence

	(1)	(2)	(3)	(4)
Tips (unincentivized)	0.12***	0.11***	0.12***	0.12***
	(0.04)	(0.04)	(0.04)	(0.04)
Tips (incentivized)	0.08*	0.08*	0.08*	0.08*
	(0.04)	(0.04)	(0.04)	(0.04)
Demographics		✓	✓	✓
Design Controls			✓	✓
Scam Experience				✓
Control Mean	4.20	4.20	4.20	4.20
p-value ($Tips^U = Tips^I$)	0.36	0.39	0.37	0.37
N	1000	997	997	991
R-Squared	0.45	0.46	0.46	0.47

Notes: Dependent variables in Panel 1 and Panel 2 are the share of correctly identified messages (SIA) and average confidence in block 2, respectively. Demographic controls include gender, age, post-secondary education, low income, formal employment), design controls include indicators for order of blocks and a dummy for attention check, controls for scam experience include low trust in DFS, above median use of different DFS, contacted less than one week ago, and victim of a scammer. All specifications include an indicator for the incentives treatment and the baseline value of the outcome variable. $Tips^U$ and $Tips^I$ refer to Tips (unincentivized) and Tips (incentivized), respectively. Displayed coefficients are from OLS regressions. Robust standard errors are in parenthesis. Asterisks indicate that the estimate is statistically significant at the 1% ***, 5% **, and 10% * levels.

C Scam Data Collection and SIA Measurement

C.1 Scam-related Keywords

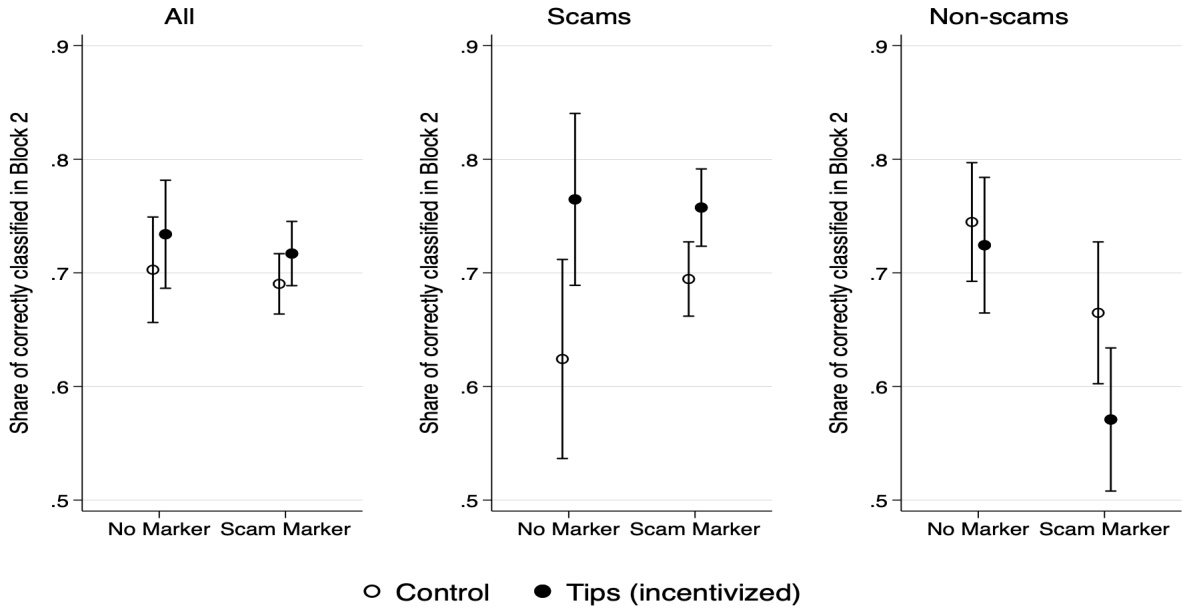
Table C1: Scam Keywords

Keyword	Description (if necessary)
Scam	
Scams	
Phone Scam	
Financial Scam	
Scammer	
Scammers	
Fraud	
Fraudster	
Digital Security	
Financial crime	
Financial fraud	
Digital Fraud	
Payment fraud	
Report fraud	
Scam call	
Scam calls	
Scam message	
Scam messages	
Fraud risk	
Tuwaanike	A campaign by Safaricom aimed at exposing fraudsters and equipping the public with information to raise the alarm on them and their tactics.
333	Safaricom's official toll-free SMS line to report fraudsters and forward their message and phone numbers for further action.
707	Safaricom's official phone number that will automatically SMS you if someone tries to register your SIM. If you reply no, it won't be registered in your place.
SIM swap	A popular scam where someone experiences their SIM card getting duplicated without their knowledge after they are socially engineered (typically over a phone call) to provide their data e.g. ID number and some more sensitive info.
"Kaa Chonjo"	A campaign by the Kenya Bankers Association aimed at equipping people with the fact that they should keep their ATM PIN a secret and stay alert at ATMs for fraudsters.
M-Pesa PIN	
Hakikisha	A tool by Safaricom to authenticate the identity of the person you're about to send money to (sometimes used to show that the fraudster uses a fake name other than the one on their ID card).
Kamiti	
Bank Account Takeover Fraud	
Debit Card Fraud	
Credit Card Fraud	
Yahoo boys	This is in reference to Nigeria scammers
Internet fraud	
mail fraud	
pyramid schemes	
identity theft	
bank card fraud	
ponzi schemes	
advance fee scam	
swift/rtgs message fraud	
forgery	
manifest fraud	This occurs when shipping agents illegally alter manifests prior to uploading them to the Customs Manifest Management System (MMS), thereby setting the stage for false declarations.
invoice fraud	
tax fraud	
phishing	
scamming	
wash wash	

Notes: Keywords were determined based on discussions with Kenyan DFS experts.

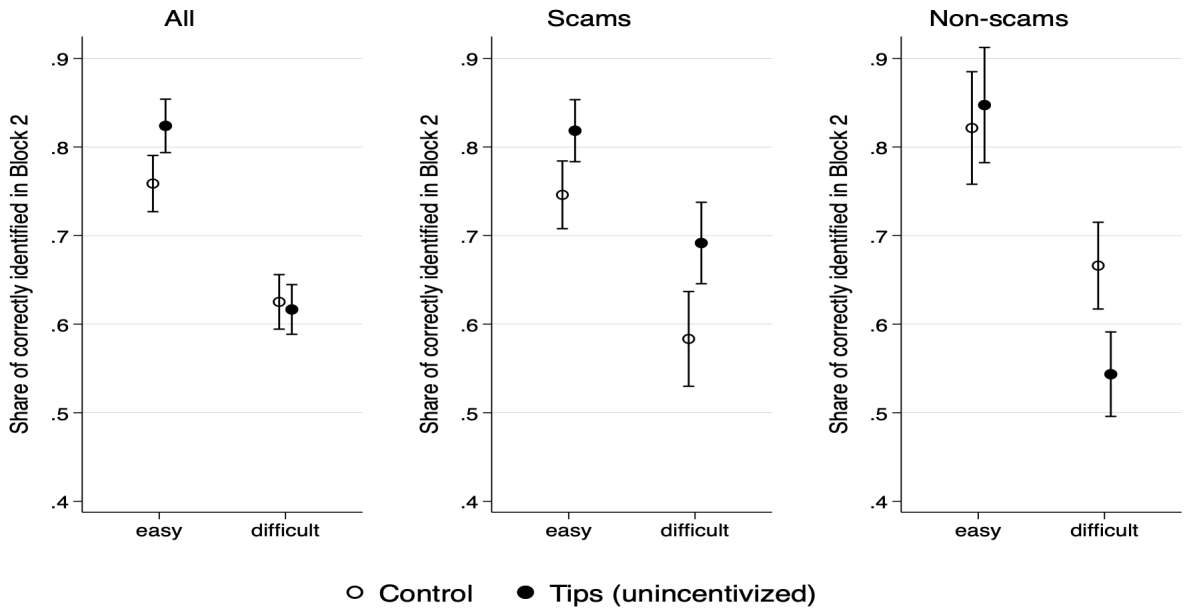
D Vignette-Level Analysis

Figure D1: Vignette-level effects by whether the message contains a scam marker



Notes: Figures plot the average marginal effects from triple-differences estimation with 95% confidence intervals based on standard errors clustered at the respondent level (see also Appendix D). Scam Marker is an indicator for whether the message contains at least one of the scam markers the tips warn about. The left panel contains all vignettes, the center panel focuses on scams, and the right panel on non scams. For ease of exposition, only the control and the Tips (incentivized) treatment are displayed. The econometric specification contains the full set of interactions and demographic controls.

Figure D2: Vignette-level effects by difficulty of the vignette



Notes: Figures plot the average marginal effects from triple-differences estimation with 95% confidence intervals based on standard errors clustered at the respondent level (see also Appendix D). Vignettes are coded as difficult if they are misclassified more often than the median vignette in block 1. The left panel contains all vignettes, the center panel focuses on scams, and the right panel on non scams. For ease of exposition, only the control and the Tips (unincentivized) treatment are displayed. The econometric specification contains the full set of interactions and demographic controls.