Standards and scoring to increase transparency for archived public opinion data

Kathleen J. Weldon[1]

Abstract

Faced with increased diversification of methodologies in the polling industry, the Roper Center for Public Opinion Research Center is embarking on a major initiative aimed at increasing methodological transparency across the field of public opinion survey research by increasing minimum disclosure requirements and providing users with transparency scoring for new submissions to the archive.

Roper Center, the world's largest archive of public opinion survey data, has long enforced disclosure requirements for archival submissions based on transparency standards developed by professional organizations in the polling industry, particularly the American Association for Public Opinion Research (AAPOR). Roper Center's new requirements and scoring mechanism expand longstanding policies and procedures to better meet the challenges of today's research environment.

In this paper, Roper Center's new standards will be described in the context of the historical development of transparency expectations in the polling community. The paper will also detail the implementation process, providing an account of how standards were translated into actionable DDI-based metadata to drive an automatic scoring system, how new workflows were developed with input from data providers to facilitate maximum disclosure, and how the display of the user interface was designed to ensure the transparency information can be easily viewed and understood.

Keywords

Transparency, polling, disclosure, public opinion, standards

Introduction

The U.S. polling community has long demonstrated a commitment to transparency, as encoded in a series of standards adopted by professional organizations in the field since the 1960s. But the rapid proliferation of new methods in polling since the turn of the century have spurred the development of more stringent and complex standards by both the National Council on Public Polls (NCPP) and the American Association for Public Opinion Research (AAPOR).

These changes in polling methodologies present unique challenges to the Roper Center for Public Opinion Research archive. The Center has long depended on methodological criteria to evaluate submissions for inclusion in the archive. The policy was described internally as preserving polling that is 'the best of its time,' in acknowledgement that methods have changed over time and the most respected polls from the 1930s preserved in the archive used quota methods that would disqualify them for inclusion in the archive if fielded in later years. In 2002, the Acquisitions

Committee of the Roper Center wrote a formal Acquisitions Policy that included the restriction that "[i]nterview data cannot be from exclusively self-selected respondents." While this wording might be interpreted in several ways, as practice the Center followed a policy of not acquiring data from online nonprobability panels.

With the rise of online nonprobability panels, interactive voice response (IVR), redirected inbound call sampling (RICS), and other new approaches to fielding surveys, the survey research community no longer shares a consensus over what constitutes best practices in the field. This change has complicated Roper's approach to acquisition. In 2018, the Board of Directors of the Roper Center approved recommendations for a new policy described in a memo from its Acquisitions and Transparency Committee. The new policy opened Roper Center's acquisitions to new methodologies, while creating a stringent set of disclosure requirements for this new collection and developing a system to score transparency across both the longstanding and recently developed collections at the Center.

This paper will trace the developments in the field of polling that led to this decision and outline the new approaches, including a description of the process of implementation.

Background: Disclosure in Polling in the 20th Century

Jane Jacobs wrote of professional self-regulation that '[a]ll variations have the self-interest of members at their core, usually sincerely construed as advancement of the profession itself.' (Jacobs, 2010, p. 128)

In the case of pollsters, self-interest might be closer to self-preservation. Unlike any other form of social science research, public opinion polls, which are frequently conducted by media organizations themselves, are released almost immediately following the completion of fieldwork, then discussed at length by media and politicians. The uniquely public role of polling has meant that from the earliest days the profession – a group that includes commercial firms, media organizations, academic research organizations, and nonprofits, with the variation in values and interests that might be expected of such a diverse group – had to invest time and effort in building the trust of politicians, journalists, and the general public. Skepticism from these groups ran high, particularly in the late 1940s when polling's massive failure to predict the winner in the 1948 Truman/Dewey race nearly destroyed confidence that had been built over the previous two presidential election success. The 1949 publication of Lindsay Rogers's *The Pollsters*, a work deeply critical of the role public opinion polling was coming to play in American life, increased the sense that this new industry was not to be trusted. Without the support of the media, and by extension the public, the field of polling could not thrive or possibly even survive.

George Gallup believed full disclosure of methods, sponsorship, and data was essential. Describing the commitment of the American Institute of Public Opinion (later the Gallup Organization) to what would come to be known as transparency, Gallup wrote:

> Since the day it was organized the American Institute of Public
> Opinion has maintained a policy of providing full information

about all of its procedures and operations. A duplicate of every ballot ever collected in its entire history is on record in the files of Princeton University for use and study by qualified students. In books, and in countless articles and speeches, we have described our methods, the size of our samples, the limitations of polls in making election forecasts, accuracy, source of revenue-which comes entirely from publications-and our overall philosophy of the place of polls in a democratic society. […] Unlike some fields, the polling profession has no trade secrets. We have held that the public has every right to know just how we function. One of the best safeguards which we have imposed upon ourselves is to report in every news release the question or questions asked, the type of cross-section (whole population over 2I, voting population, informed public, etc.), along with the results. (Gallup, 1948)

Gallup's admirable openness was, as he notes, 'self-imposed.' Prominent pollsters, Gallup included, had discussed the potential value of setting professional standards for reporting from their first meetings together. As described by Sidney Hollander in *A Meeting Place*, the history of the American Association for Public Opinion Research (AAPOR), the idea of establishing a set of reporting standards was raised at the Central City conference in 1946, the precursor to AAPOR's yearly conference (Hollander, 1992). In 1948, at the meeting where the AAPOR's constitution was written, a set of disclosure standards was also drafted, though no action was taken to move forward with the adoption. The debate over standards continued without action for twenty years.

In 1967, AAPOR finally made its move. Gallup led the charge, concerned that the field was threatened by a proliferation of bad actors using questionable methods and, particularly in the case of the rapidly expanding field of political polling, releasing partial results intended more to influence than to reflect public opinion (Gollin, 1992). A set of disclosure standards was adopted by AAPOR Council.

Another form of pressure had surely influenced this decision. The specter of government regulation that had long hung over the industry had grown more threatening. In 1943, Senator Gerald Nye had proposed a bill that would have required pollsters to disclose sample size and retain records for two years. No action was taken, but the warning bell had been rung. In 1968, as the AAPOR membership was first learning of the new standards Council had committed to the previous year, Rep. Lucien Nedzi of Michigan sponsored a bill with real teeth. His legislation set disclosure standards to be enforceable by a fine of $1000 or 90 days in jail or both. His required items for reporting looked similar to the list first suggested in 1948, covering sponsorship and basic methodological details. In an article in *Public Opinion Quarterly*, Rep. Nedzi directly addressed the polling community, suggesting the 'prospect of legislation' might be as effective as legislation itself, motivating pollsters to self-police. (Nedzi, 1971)

In 1979, the AAPOR standards served as the basis for a new set of standards adopted by the National Council of Public Polls (NCPP). Over the next few decades, major polls published results with a

methodology statement followed by what became a familiar notation: 'These statements conform to the principles of disclosure of the National Council on Public Polls.'

Peer-to-Peer Transparency

Although Gallup had boasted of his submission of data punch cards to a repository at Princeton (later moved to the Roper Center), as well as conference presentations and published articles, for decades the primary focus of all debates over disclosure had been public reporting, not data sharing. The audience of concern was the media, and by extension government and the people. In his chapter in *A Meeting Place*, Albert Gollin described the concerns about disclosure that led to the first official standards in the 1960s as 'a struggle about control over the release of public opinion data to the public as well as about how to educate the press and public concerning the hallmarks of a professionally conducted survey.' (Gollin, 1992) The focus on the media continued into the professional literature on standards. In a 1982 *Public Opinion Quarterly (POQ)* article, Miller and Hurd noted that the AAPOR and NCPP polls were primarily intended to provide disclosure guidelines for survey researchers in releasing polls, but also that 'it is obvious they were also meant to sensitize journalists.' (Miller & Hurd, 1982) A number of academic articles over the 1980s and 1990s attempted to measure the success of the NCPP standards by determining what proportion of media reports on polls included the required information. Implementation of disclosure in media reporting on polls was also the topic of 1971 and 1980 *Public Opinion Quarterly* symposiums and a 1979 NCPP/Kettering Foundation conference.

Data sharing or requirements intended to explicate methodology at a level of detail required for researcher analysis were not part of the discussion. Sharing of methodological information among polling professionals continued just as Gallup described, through annual AAPOR meetings and other conferences, in ad hoc AAPOR committees, in the pages of *POQ* and other academic journals, and at the Roper Center archive, which maintained a minimum disclosure requirement for acquisition that closely followed the NCPP and AAPOR standards.

In 2006, everything changed. The National Council of Public Polls created an expanded three-level disclosure standard. (NCPP, n.d.) Level one concentrated on the traditional information required with public release of results. The second level focused on information that member organizations had to make available upon written request. The items in this level were far more comprehensive that those at the first level. The third level, which was strongly encouraged, but not required, was the release of datasets. The intended audience for these additional layers of requirements were clearly other members of the polling community. Even the most poll-savvy reporters or citizens were not expected to make judgments about weighting methods or disposition codes, much less to wrangle SPSS files.

In 2008, the AAPOR community found evidence that lack of transparency was preventing the field from identifying the problems that had plagued that year's primary election polling. The willingness of polling organizations to share detailed methodological information and datasets had helped the industry overcome its failures in the 1948 election. But the report of the Ad Hoc Committee on the 2008 Presidential Primary Polling repeatedly noted the failure of survey organizations to provide

timely and thorough methodological information. (Traugott et al., 2009) Twenty-one organizations provided at least some information, but three organizations never responded to the Committee's request for data at all. While the majority of responding organizations provided information on weighting and question wording, only seven provided the microdata, which was then deposited at the Roper Center. Just four fulfilled the request for data on the gender and race of interviewers. As a result of these omissions, the Committee called for a review of disclosure standards.

In 2010 AAPOR announced the establishment of the Transparency Initiative (TI), creating a membership program which polling organizations could join by committing to abiding by the new disclosure standards. Like the NCPP standards, AAPOR included a set of additional disclosure items to be made available upon request. In a July 2012 presentation at the RC-33 Conference, TI Committee Chair Timothy Johnson and Paul Lavrakas identified the primary problem as 'inadequate transparency of research methods and statistical methods' that causes a 'serious detriment to progress.' (Johnson & Lavakras, 2012) The main goal of the initiative was to 'advance the science and reputation of survey research', while public education on transparency was secondary. Although neither archiving nor sharing of the dataset was required in the standards, by establishing a much greater level of transparency expectation upon request, AAPOR expanded its focus on disclosure from journalists and the public to peer-to-peer transparency.

Evolving needs

Why did NCPP and AAPOR both increase their requirements so dramatically within in a few short years? The movement toward greater transparency in polling was part of a larger shift toward new expectations of data sharing, replication, and transparency in social science research, as described by Herndon and O'Reilly in 2016. New requirements were enforced by journals in which polling researchers often publish, like the *American Journal of Political Science,* which adopted data sharing requirements in 2012, as well as funding agencies that support academic pollsters, like the National Science Foundation, which incorporated data sharing requirements into large grants in 2011, and NIH, which did so even earlier, in 2003. (AJPS, n.d.; NSF, n.d.) But polling as a discipline had another driver towards greater transparency. Polling returned to the question of transparency standards in the early 2000s when several new and controversial methods, most notably internet panels, began to become mainstream.

In a 2005 article, Mark Blumenthal built a case for increased transparency in polling by tracing recent increases in methodological heterogeneity.(Blumenthal, 2005) The first of the internet opt-in panel pollsters, Harris Interactive, had conducted polls during the 2000 election, but in the next presidential cycle, multiple organizations jumped into the new methods sphere, with online panel pollsters Zogby International and British firm YouGov, and interactive voice response (IVR, or 'robocall') pollsters SurveyUSA and Rasmussen drawing major media attention and enormous Internet traffic. Blumenthal argued that, not only these new methods, but new dissemination approaches changed the polling landscape during the first decade of the new millennium, as some new polling organizations began to publish their results directly on their own websites, rather than through major media outlets. These organizations were able to receive wide attention for their polls without undergoing the standard vetting process used by most major media organizations. In

response to these developments, Blumenthal called upon survey researchers to embrace transparency of methodology, specifically citing rapid rate of change as the primary reason for increased need for disclosure.

Since 2005, polls based on recently developed methods like IVR and online nonprobability panels have increasingly entered the mainstream, despite ongoing concerns about data quality and accuracy.[2] *The New York Times* and *The Economist* both have partnered with YouGov, *Washington Post* and *Business Insider* with SurveyMonkey, and *USA Today* with Ipsos Public Affairs, all utilizing online non-probability panel methods. These approaches are also expanding the types of populations polled. New organizations have been taking advantage of the lower costs of targeting historically under- polled groups using new methods by developing polling projects focused on these populations, such as Latino Decisions, Asian American Decisions, the African American Research Collaborative, and the American Muslim Poll.

The new AAPOR disclosure standards included a number of items aimed specifically at new methodologies, including disclosure of use of routers (sites that connect potential respondents with online surveys for which they are eligible) and specific recommendations for the reporting of sampling error estimates in nonprobability polls. When AAPOR announced its new standards, Republican pollster David Hill wrote approvingly of the effort in *The Hill*, tying the need for new standards directly to the explosion of new methods: 'as data collection methods and sampling frames have become more exotic, including robo-calls and online panel surveys, new standards are clearly indicated.' (Hill, 2010) The relationship of new methods and disclosure was also apparent in the report of the Ad Hoc Committee on the 2008 Presidential Primary Polling, which in calling for a review of disclosure standards specifically referenced the new world of 'more complicated and diverse sampling frames and selection techniques' and 'more complicated and diverse statistical adjustments for errors of non-observation.' (Traugott et al, 2009)

Roper Center's Transparency Project: New Standards

The proliferation of new methods polling presented a challenge to the Roper Center's traditional approach to collection. The acquisitions policy, adopted in 2002 and most recently reviewed in 2012, specified the use of probability-based methods, while the use of IVR technologies and voter file samples were not specifically prohibited, but in practice had been avoided in collection. The policy also specified required elements of disclosure reflective of the standards of NCPP and AAPOR before their revisions. The field of polling research had changed, and the Roper Center had to respond thoughtfully. The need to accurately represent current methods had to be balanced with the Center's reputation as an archive that preserved 'the best of its time.' The Center also had to weigh increased expectations of disclosure with a commitment to maintain overall transparency in the field by ensuring strict new requirements did not cause current donor organizations to stop sharing data.

In June 2018, after several years of deliberation, the Acquisitions and Transparency Committee of the Roper Center's Board of Directors submitted a memo to the full Board proposing a bold new Transparency Project with two major initiatives: a transparency scoring metric to be displayed on all new dataset catalog entries and the establishment of a new collection of surveys conducted using

recently developed methods. The new collection would be open to all methodologies, allowing researchers to analyze these methods and potentially improve upon them. However, in recognition of concerns about possible data quality issues, several conditions would apply. All recently developed methods submissions would need to meet a high bar of transparency. Only questions with dataset submissions would be included in the new methods database, in contrast to the longstanding methods collection, which includes topline results without underlying data files. Finally, the new collection would be searched and displayed separately from the longstanding methods collection on the Roper Center website. These safeguards were particularly important for Roper Center users who are not advanced researchers in the field, a group that includes undergraduates and some media and nonprofit users. The Board approved these recommendations.

Transparency Scoring

The scoring system groups disclosure elements into 'Core' and 'Additional.' Core items will be required for all recently developed methods submissions and strongly encouraged for longstanding methods studies. In order to ensure that overall transparency in the field was not reduced by a sudden increase in requirements that might lead longtime data providers to stop sharing data with the Center, the Committee decided not to change existing requirements for traditional methods studies. Over time, the Center hopes that the Transparency Project will provide an incentive for all data providers to adopt disclosure of the Core items as recognized best practice.

This scoring system was heavily influenced by the AAPOR and NCPP standards, and overlap across the different standards is significant. Figure A, which builds upon work by Lois Timms-Ferrara and Marc Maynard, provides an overview of the elements included in each of the major proposed and enacted standards from 1968 to today, showing how standards have grown in scope and complexity.(Timms- Ferrara & Maynard, 2011) Empty cells represent items that are not included in a standard. In recent years, more disclosure has been required or encouraged and more focus has been brought to bear on questions of weighting and sampling, both essential in understanding new methods.

	Nedzi proposal	AAPOR 1967	NCPP (pre2006)	NCPP (Current)	AAPOR (Current)	Roper
Summary information						
Survey field organization				Level 1	Immediate	Core
Sponsor/funder	X	X	X	Level 1	Immediate	Core
Population		X	X	Level 1	Immediate	Core
Dates of interviewing	X		X	Level 1	Immediate	Core
Timing of interviewing in relation to events		X				
Topline results	X		X	Level 1		Core*
Sample size	X	X	X	Level 1	Immediate	Core
Size of any subgroup included in the report		X	X	Level 1	Immediate	Core*
Weighted and unweighted				Level 2		Core*
Margin of error				Level 1	Immediate	

7/19   Weldon, Kathleen J. (2020) Standards and scoring to increase transparency for archived public opinion data, IASSIST Quarterly 44(3), pp. 1-19. DOI: https://doi.org/10.29173/iq974

Other description of estimated accuracy					Immediate	
Questionnaire/Instrument						
Exact wording of questions/responses	X	X	X	Level 1	Immediate	Core
Exact wording of introduction				Level 2		Core
Interviewer or respondent instructions					Within 30 days	Core
Languages in which survey was offered					Immediate	Core
Complete wording of questions in any foreign languages in which the survey was conducted				Level 2		
Any relevant stimuli/visual aids					Within 30 days	Core
Sampling						
Sampling method	X			Level 1	Immediate	Core
Margin of sampling error		X		Level 1	Immediate	
Whether these have been adjusted for design effect due to weighting, clustering, or other factors					Immediate	
Justification for claims of representativeness						Core
Coverage of target population/Estimated size of the noncovered population				Level 2	Immediate	Additional
Sample design/sampling frame(s)					Immediate	Core
Name of the sample supplier, if sample/frame provided by third party					Immediate	Additional
Proportion of sample provided						Additional
The methods used to recruit the panel or participants, if applicable					Immediate	
Respondent selection procedure (for example, within household), if any				Level 2	Immediate	Core
Description of any quotas or additional sample selection criteria during or post fielding					Immediate	
Maximum number of attempts to reach respondent				Level 2		
Incentives					Within 30 days	
Other strategies to gain cooperation					Within 30 days	
Use of breakout routers or chains					Within 30 days	Additional
Details about other types of screening procedures					Within 30 days	
Interviewing						
Method of interviewing (mode)	X	X	X		Immediate	Core
Response rates	X				Within 30 days	Core**
Completion or participation rate (surveys for which a response rate cannot be calculated)						Core**
Sample dispositions adequate to compute contact, cooperation and response rates				Level 2	Within 30 days	Core**

Minimum number of completed questions to qualify a completed interview				Level 2		
Breakoff rate						Additional
Whether interviewers were paid				Level 2		
Incentives or compensation provided for participation				Level 2	Within 30 days	Additional
Weighting						
Description of weighting procedures (if any) used to generalize data to the full population				Level 2	Immediate	
Weighting benchmark source					Immediate	Core
Variables used to calculate weights					Immediate	Core
Identification of weighting variable in dataset						Core
Quality control						
Procedures for managing the membership, participation, and attrition of the panel, if applicable					Within 30 days	
Methods of interviewer training, supervision, and monitoring, if interviewers were used				Level 2	Within 30 days	
Quality control procedures/data verification					Within 30 days	Additional
% respondents removed due to quality control checks						Additional
Datasets						
Release deidentified raw datasets				Level 3		Core
Operational and reporting						
Post complete wording, ordering and percentage results of all publicly released survey questions to a publicly available web site for a minimum of two weeks				Level 3		
Publicly note their compliance with these Principles of Disclosure				Level 3		
Contact name and information					Immediate	
Survey organizations reporting results will endeavor to have print and broadcast media include the above items in their news stories and make a report containing these items available to the public upon release			X			
Specifications adequate for replication of indices or statistical modeling included in research reports.			X		Within 30 days	
Wordings to describe similar disclosure items vary across different standards. Please see individual standards for complete wordings.						
*This information can be derived from the dataset, a Core Roper Center item.						
**Roper Center considers either response rate and AAPOR definition or disposition codes to calculate the same sufficient to meet Core disclolsure.						

Implementation of Scoring

In order to implement the Committee's recommendations, Roper staff had first to map the elements the Committee had identified to existing DDI-based metadata in the Roper database. In many cases, the mapping was a simple one-to-one connection to existing metadata elements. In some cases, however, a single element from the Committee recommendations actually represented several metadata fields, as described in the DDI standard. For example, the list of 'modes' as described by the Committee included the concepts of both sampling procedure and mode. Some Committee recommendations expanded the number of metadata fields that the Roper Center will need to capture. The number of fields related to weighting, for example, has increased from one to four, three required by the new system and an additional notes field.

After the full range of necessary disclosure items had been defined by Roper staff and verified by the Committee, the staff then had to determine the type of metadata field needed. In most cases, the elements of transparency scoring consisted of fields that could be entered in numerical or text formats. But in some cases, the element represented an indicator of whether certain material was included in the archival package: for example, visual aids or complete interviewer instructions. In these cases, the staff determined the most practical approach to be a simple checkbox to indicate presence of this item in the downloadable documentation file.

Staff also reviewed each element to determine if in any case it might be inapplicable to a particular survey based on methodology or other reasons. For those items, a 'not applicable' option would need to be available to avoid surveys being docked points for 'missing' inapplicable information in an autoscoring system. Finally, staff wrote definitions for each element to ensure that each item was understandable and clear to both data providers and end users.

Field	Definition	Acquisition Committee Memo Item	NA option	Field type
Survey sponsor	When applicable, the name of the organization that commissioned the survey. If the same organization funded, designed, and fielded a poll, no sponsor is listed.	Survey sponsor, including all funding sources	Yes	Open text
Grant funding source	Funding source for academic or other grant-supported research.	Survey sponsor, including all funding sources	Yes	Open text
Survey organization	The organization that conducted the fieldwork for a survey.	Field work provider, if outsourced	No	Open text
Data collection dates	The date range during which data was collected from respondents.	Interview dates	No	Date

Universe	The population the survey results are intended to represent. Also known as "target population."	The population of which the results are said to be representative, and the justification for this research claim, AND The universe from which the sample was drawn, and the proportion of that universe that had a nonzero chance of participation	No	Open text
Geographic coverage	The geographic area from which data were collected.		No	List
Justification for claims of representativeness	A description of the elements of the research design intended to ensure that the survey is representative of the universe it is designed to study.	The population of which the results are said to be representative, and the justification for this research claim	No	Open text
Mode	Method by which data were collected (such as telephone, in-person, online, etc.)	Mode: RDD telephone, IVR; listed-sample telephone with live interviewers; listed-sample telephone via IVR; other telephone (describe); opt-in online panel; other online (e.g., river samples, mobile apps; hybrid or other (describe))	Yes	List
Mode other: Description (filtered on previous)	Method by which data were collected, such as telephone, in-person, online, etc.		Yes	Open text
Sample size	The total unweighted number of respondents in the survey.	Unweighted sample size	No	Numerical
Sampling procedure: Summary	The method by which participants in a poll were selected.	Sampling method: Probability, non-probability or hybrid AND Mode: RDD telephone, IVR; listed-sample telephone with live interviewers; listed-sample telephone via IVR; other telephone (describe); opt-in online panel; other online (e.g., river samples, mobile apps; hybrid or other (describe))	No	Open text
Sampling procedure: Respondent selection stage	The method by which participants in a poll were selected; specifically, the method by which the individual respondents were chosen. In a multistage sampling process, respondent selection is the final stage of sampling.	Respondent selection procedure, or absence thereof	No	Open text

Sampling frame	A list of the items or people forming the universe from which a sample is taken.	Sample frame and a description of the universe from which the sample was drawn Description of all sample weights and sources of weighting targets	No	Open text
Weight variable	Name of the variable in the datasets used for weighting the sample. If mutiple weighting schemes were used for different analysis, the variable identified here will be the one used for reporting on the total population, and information on other weights provided in the documentation.		Yes	Open text
Weighting benchmark source	Data source for benchmarks used to weight the sample		Yes	Open text
Variables used for weighting	Specific variables used in the calculation of survey weights.	Response rate calculated to AAPOR standards, or sample disposition data adequate for the calculation of AAPOR- standard response rates. When AAPOR- standard response rates cannot be calculated, completion or participation rates shall be provided using another method that is fully disclosed Response rate calculated to AAPOR standards, or sample disposition data adequate for the calculation of AAPOR- standard response rates. When AAPOR- standard response rates cannot be calculated, completion or participation rates shall be provided using another method that is fully disclosed	Yes	Open text

Response rate*	Proportion of contacted respondents who completed the survey. The American Association for Public Opinion Research (AAPOR) provides definitions for six measures of response rates.		Yes	Numerical
Disposition codes*	A set of codes or categories used by survey researchers to document the ultimate outcome of contact attempts on individual cases in a survey sample.		Yes	Checkbox
Completion or participation rate	The proportion of all cases interviewed of all eligible units ever contacted, used if response rates calculated to AAPOR standards would be inappropriate for the survey design.		Yes	Numerical
Completion or participation rate details (filter on previous)	Method for calculation of completion/participation rates for surveys for which standard AAPOR response rates cannot be calculated	Survey language(s)	Yes	Open text
Survey language(s)	Languages in which the survey was fielded.	Survey language(s)	No	List
Full question wording with all interviewer instructions, prompts and visual aids	A complete survey questionnaire includes all questions, including any screening questions, introductory language, interviewer instructions, and, in the case of some in-person or online polls, visual aids used to illustrate questions.	Full survey questionnaire with all instructions, prompts, visual aids	No	Checkbox

External sample provider(s)	The organization that provided the sampling frame to the field organization, if external sample provider used.	Sample provider(s), and, if multiple, the share of sample from each provider	Yes	Open text
Proportion of sample provided (filtered on previous)	The proportion of the total sample provided by the external sample provider.		No	Numerical
Use of breakout routers or chains	Use of online survey routers that screen respondents and direct them to open surveys for which they are qualified or use of chains that direct respondents to additional surveys at the end of completed surveys.	Use of survey routers or chains	Yes	Checkbox
Breakoff rate	The percent of respondents who start the survey but do not finish it.	Breakoff rate (i.e., the percent of respondents who start the survey but do not finish it)	No	Numerical
Estimated size of the noncovered population	Proportion of universe that had a nonzero chance of participation	The universe from which the sample was drawn, and the proportion of that universe that had a nonzero chance of participation Use of incentives	No	Numerical
Use of incentives	Use of incentives provided to survey recipients to reward participation.	Use of incentives	No	Yes/No

What incentive was provided (filter on previous)	Specific incentives provided to survey recipients to reward participation.		Yes	Open text
Quality control checks	Quality control checks performed on the data from the survey. Many possible approaches can be taken for quality assurance, such as monitoring online surveys for cases of "speeding" (answering at a rate too fast to allow for adequate comprehension of questions) or "straightlining" (providing identical answers across a range of questions); reinterviewing in-person survey respondents; or random quality control monitoring of telephone interviews.	Details of quality control checks (e.g., for logic, speeding, straightlining), including how they were performed and results of those checks, including percent of completed interviews excluded or dropped from the analysis Details of quality control checks (e.g., for logic, speeding, straightlining), including	No	Checkbox
% respondents removed due to checks (filtered on above)	Percentage of respondents whose cases were removed from the survey before analysis based on quality checks performed.		Yes	Numerical

At this point, the Center data staff shared back with the Committee the translation of their work into a plan for a functional scoring system, and after some collaborative revision, the list of elements on which scoring would be based was finalized.

The next task was to score one study from the most recent submission from each of thirty-two active data providers. New questions emerged as a result of applying the scoring mechanism to actual studies and were presented by staff to the Committee for review. Could a survey that was fielded on an omnibus be considered to include 'all question wordings, including interviewer instructions' when the sponsor was unable to provide the introductory text and other questions asked on the same instrument? (Yes.) Is an average response rate for a tracking poll sufficient when a data provider submits a monthly aggregate of daily polls? (Yes.) If a multicountry poll offers

different levels of disclosure for different countries, should the highest or lowest level of disclosure be used in scoring? (Lowest.) How much information on respondent selection method in an RDD survey is needed to satisfy the requirement? ('Random' is not enough; method of randomization must be provided.) Although refining standards will of course be an ongoing process, this initial effort by Center staff in collaboration with the Committee ensured that procedures for dealing with the most common issues were in place.

Display and Design

With development of the scoring system underway, the Center staff and Committee also considered the issue of display. To lead users to meaningful engagement with the scoring system, only a button reading 'Transparency Details' will show on search results pages. This button will lead to a page on which a numerical score will be provided, based on the following formula:

((10 points for providing a dataset + 2 points for every other applicable core item + 1 point for every applicable additional item))/(total possible points for all applicable items)) X 10. (Results rounded to the nearest .5)

Studies will also be assigned to one of three descriptive categories. Studies with a score >=9 and <=10 'Greatly Exceed Requirements;' scores >=8 and <9 'Exceed Requirements;' and scores >=6 and <8 'Meet Requirements.' No study meeting current acquisitions guidelines could score below a 6. These categories were chosen by the Committee to frame scoring appropriately: any data provider to the Center meets a high standard of transparency, and scoring simply expands upon that baseline. However, the categories are also intended to offer data providers an incentive for offer more information. Under the category and numerical score, the elements are provided in a checklist to offer users a quick overview of available documentation.

Into the Future

At the time of writing, the Center is conducting outreach to current and potential data providers to describe and explain the Transparency Project; some changes may result from this effort. The Center is also inviting feedback from the broader polling research and data archives communities, both now and in the coming years as this project evolves to reflect the rapidly changing polling research environment. Development of automated scoring in the ingest system is ongoing. After the completion of that project, scoring will be integrated into the member website. At that point, the display of an overview of available information will be a convenience that should aid research. If the effort is also successful in increasing the information provided by data providers, the opportunity to judge data quality and compare the effects of different methodological approaches should increase.

But will the Transparency Project actually increase transparency? As with so many such questions, it may depend on how success is measured. The results of the AAPOR TI to date have been hard to quantify. The TI boosts an impressive list of nearly ninety members. However, formal requests for additional information, which are channeled through the AAPOR Transparency Committee, have been few and far between. During the tenure of the first chair of the Transparency Committee, no

request was made, and only three have come through since the second chair took over.(Johnson, T. personal correspondence, February 20, 2019; Kirzinger, A., personal correspondence, February 11, 2019) Informal requests sent directly to survey organizations by researchers, however, are not recorded, and therefore the standards may have had an impact that is currently undocumented.

After the 2016 election, AAPOR once again appointed a committee to review failures in state-level election polling. The committee contacted 59 organizations, an increase over 2008 that likely reflects both the broader geographical scope of the committee's charge and the proliferation of polling operations. Only 35 responded. (Kennedy et al, 2018) This low response rate seems to indicate that the TI's hopes of increasing transparency in polling have not been fulfilled. However, none of the non- responders was part of the TI. (Kennedy et al, n.d.)

The polling industry may be moving in two directions at once. As more and more inexpensive online polls are conducted by new organizations with, as Mark Blumenthal noted back in 2010, little connection to professional associations in the field and no need to rely on the vetting process of major media outlets for dissemination, the overall level of disclosure in the field may decrease. However, those organizations that have embraced the polling evolving commitment to transparency may, under the influence of the NCPP standards, the AAPOR TI, and Roper Center's Transparency Project, provide far more comprehensive information in both their initial releases and in their archival submissions. The body of well-documented polling datasets preserved for the future should increase substantially.

Currently non-archiving organizations, in choosing to align themselves with the 'transparency-committed' sector of the polling world, may decide to share their data through the Roper Center and ensure researcher access to more data now and into the future. These results would represent a major success not only for the Roper Center's Transparency Project, but for the field of public opinion polling as a whole.

References

AAPOR, November 2105, *Code of Ethics 2015*, viewed 20 December 2019. https://www.aapor.org/Standards-Ethics/AAPOR-Code-of-Ethics.aspx

American Journal of Political Science, (n.d.), *AJPS Replication and Verification Policy*, viewed 20 December 2019. https://ajps.org/ajps-verification-policy/

National Council on Public Polls, (n.d.), *Principles of Disclosure*, viewed 20 December 2019. http://www.ncpp.org/?q=node/19

Asher, H 2001. *Polling and the public: What every citizen should know (5th edn.)*, CQ Press, Washington, DC.

Blumenthal, MM 2005, 'Toward an open-source methodology: What we can learn from the blogosphere', *Public Opinion Quarterly*, vol. 69, no. 5, pp. 655–669. https://doi.org/10.1093/poq/nfi059

Gallup, G 1948, 'On the regulation of polling', *Public Opinion Quarterly*, vol. 12, no. 4, 733–735. https://doi.org/10.1086/266023

Gollin, AE 1992 'AAPOR and the media', in PB Sheatsley & WJ Mitofsky (eds.) 1992. *A meeting place: The history of the American Association for Public Opinion Research*, American Association for Public Opinion Research.

Herndon, J, & O'Reilly, R. (2016). Data sharing policies in social sciences academic journals: Evolving expectations of data sharing as a form of scholarly communication. *Databrarianship: The academic data librarian in theory and practice*, 22.

Hill, D 2010, 'AAPOR updates poll standards,' *The Hill,* May 18.

Hollander, S 1992 'Survey standards', in PB Sheatsley & WJ Mitofsky (eds.), 1992. *A meeting place: The history of the American Association for Public Opinion Research*, American Association for Public Opinion Research.

Jacobs, J 2010. *Dark age ahead.* Vintage Canada, Toronto.

Johnson, T & Lakavrakas, P 2012, 'AAPOR's Transparency Initiative,' paper presented at RC-33 *Eighth International Conference on Social Science Methodology*, Sydney, Australia, viewed 20 December 2019. https://slideplayer.com/slide/6530738/

Kennedy, C, Blumenthal, M, Clement, S, Clinton, JD 2018. 'An evaluation of the 2016 election polls in the United States,' *Public Opinion Quarterly*, vol. 82, no. 1, pp. 1-33. https://doi.org/10.1093/poq/nfx047

Kennedy, C, Blumenthal, M, Clement, S, Clinton, JD (n.d.), 'An evaluation of the 2016 election polls in the United States', American Association for Public Opinion Research, viewed 20 December 2019, https://www.aapor.org/Education-Resources/Reports/An-Evaluation-of-2016-Election-Polls-in-the-U-S.aspx.

NCPP (n.d.), *Principles of Disclosure*, viewed 20 December 2019. http://www.ncpp.org/?q=node/19

Maynard, M & Timms-Ferrara, L 2011, 'Methodological disclosure issues and opinion data,' *Journal of Economic and Social Measurement*, vol. 36, no. 1-2, pp. 19-32. https://doi.org/10.3233/JEM-2011-0340

Meyer, P 1968, 'Truth in polling,' *Columbia Journalism Review*, vol. *7*, no. 2, p. 20.

Miller, MM, & Hurd, R 1982, 'Conformity to AAPOR standards in newspaper reporting of public opinion polls,' *Public Opinion Quarterly*, vol. 46, no. 2, pp. 243-249. https://doi.org/10.1086/268716

Nedzi, LN 1971, 'Public opinion polls: Will legislation help?' *The Public Opinion Quarterly*, vol. 35, no. 3, pp. 336- 341. https://doi.org/10.1086/267917

Traugott, M, Bolger, G, Davis, DW, Franklin, C, 2009, 'An evaluation of the methodology of the 2008 pre-election primary polls,' American Association for Public Opinion Research, viewed 20 December 2019, https://www.aapor.org/Education-Resources/Reports/Methodology-2008-Primary-Polls.aspx

End-notes

[1] Kathleen J. Weldon is the Director of Data Operations and Communications at the Roper Center for Public Opinion Research, Cornell University. She can be reached at kjw93@cornell.edu.

[2] For an in-depth exploration of the problems with nonprobability polling, see MacInnis, B, Krosnick, JA, Ho, A, & Cho, MJ 2018. 'The Accuracy of Measurements with Probability and Nonprobability Survey Samples: Replication and Extension', *Public Opinion Quarterly.*