The Interviewer Performance Profile (IPP): A Paradata-Driven Tool for Monitoring and Managing Interviewer Performance

Survey Methods: Insights from the Field

Heidi M. Guyer | Brady T. West | Wen Chang

How to cite this article : Guyer, H.M., West B.T. & Chang W. The Interviewer Performance Profile (IPP): A Paradata-Driven Tool for Monitoring and Managing Interviewer Performance. Survey Methods: Insights from the Field. Retrieved from https://surveyinsights.org/?p=15306

DOI : 10.13094/SMIF-2021-00005

Copyright : © the authors 2021. This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0)

Abstract : Monitoring interviewer performance during data collection is essential for ensuring optimal performance at the interviewer level, thereby leading to improved data collection parameters in areas such as effort and efficiency. Paradata are widely utilized to enhance typical measures of performance, such as response rates, and provide a more nuanced view of interviewer performance characteristics. We describe a paradata tool developed to monitor interviewer performance in real time: the Interviewer Performance Profile (IPP). Daily data updates allow for ongoing monitoring to detect areas of performance concern as well as areas of improvement over time. Additionally, the flexible nature of the IPP facilitates adaptation by other projects. An overview of the tool is provided in addition to several examples of implementation and usage in a national field project.

Introduction

The use of paradata to monitor data collection outcomes has become a norm in survey research. Paradata have been used to monitor interviewer performance during interviewer training as well as during data collection. Paradata are also frequently used in static reports and production monitoring dashboards to quickly determine whether data collection parameters are on track or in need of intervention (Groves and Heeringa, 2006). There are many benefits and goals of these types of dashboards, including improving efficiency, identifying areas of concern and improvement, identifying poor performers, minimizing non-response error, reducing or controlling costs, and achieving the study outcomes. Paradata monitoring tools can be utilized by field managers as well as research staff and clients, providing the same information to all team members to ensure productive discussions, prompt decision-making, and improvement of data collection outcomes. The use of paradata dashboards is considered a best practice by many survey methodologists (Mohadjer and Edwards, 2018).

The focus of this report is on paradata-driven tools for monitoring and improving interviewer performance. Such tools have evolved over the past several decades and have been utilized to assess various types of performance. Groves and McGonagle (2001) utilized interviewer training data to develop a tool to predict future cooperation rates of individual interviewers. Durand (2005, 2008) developed the “Net Contribution

to Performance Index (NCPI)", in which interviewers received a score that took the difficulty of the task being measured into consideration. Laflamme and St-Jean (2011) further developed the NCPI by including "objective interviewer performance indicator that takes into account the complexity of the survey data process." This progression demonstrated the need to account for project and task difficulty when measuring interviewer performance. West and Groves (2013) continued this line of work and developed a novel measure of interviewer performance taking into account the difficulty of the assigned sample or achieving the outcome itself to compute an interviewer-specific Propensity-Adjusted Interviewer Performance (PAIP) Score. PAIP scores can be used to evaluate their difficulty-adjusted effectiveness in terms of various key parameters, including contact rates, response rates, and eligibility rates.

The progression of tools for monitoring interviewer performance has been accompanied by an increase in the number of variables being measured. Early on, Sudman (1967) found that simpler reporting approaches worked best. Field supervisors are the first in line to assess interviewer performance and tend to be the most accurate people when evaluating interviewers' quality over time. The EPIC-German study evaluated interviewer performance using in-depth qualitative and quantitative measurements of quality control from direct observations which were then coded into performance scores (Kroke, Bergmann, Lotze, Jerckel, Klipstein-Grobusch, and Boeing, 1999). Guyer and Wagner (Guyer 2004; Wagner and Guyer, 2005) found that statistical processing reports with highlighting to indicate poor performers versus good performers were useful visual tools for managers. Work conducted by Olson and Bilgen (2011) suggests that many typical indicators of good interviewer performance, such as faster interview pace, may lead to higher levels of acquiescence from respondents than the pace typically exhibited in a less experienced interviewer, or one that has longer-running interviews.

In this paper, we describe a robust and dynamic tool for monitoring interviewer performance over time that builds on these prior efforts and provide specific examples of its use on a national field study.

Methods

The National Survey of Family Growth's Interviewer Performance Profile (IPP)

The IPP was developed for a national survey of reproductive health, the National Survey of Family Growth (NSFG). The NSFG is a cross-sectional population-based survey on factors affecting pregnancy and childbearing including contraception, sexual activity, marriage, divorce and cohabitation, as well as services relating to pregnancy, infertility and reproductive health in the United States. A national probability sample frame is used to select approximately 5,200 addresses on a quarterly basis. Sampled households are screened and an age-eligible (15-49 years old) household member is selected and invited to complete an interview. All interviews are conducted in-person by female interviewers. Given the continuous nature of the data collection (2006-2010, 2011-2019), data exists to compare current performance to previous quarters, years and cycles of data collection. Approximately 45 interviewers are employed on the study each year. Annual trainings take place for newly hired interviewers whereas existing interviewers simply continue from one year to the next if their geographic area is in the study in sequential years. Each year, approximately 40-45% of the interviewers are new to the project and to the survey organization. The IPP allows the project team to monitor interviewer performance on a daily basis among both new and experienced interviewers and identify specific areas where additional training or guidance may be necessary in order to improve performance. Additionally, post-training changes in performance can be tracked in the IPP.

A birds-eye view of the IPP is provided in Figure 1. The four main areas of performance are indicated, with one column for each parameter evaluated and one row per interviewer. This visual display allows for both a quick visual evaluation of interviewers that are performing well or performing poorly for any given indicator (e.g., main interview response rates, conditional on completed screening interviews) as well as to evaluate interviewers that are performing well in numerous areas versus those that are struggling in numerous areas. In addition to quickly evaluating interviewer or parameter performance, an overall indication of the key areas where additional reinforcement may be needed are easily identified. The IPP is a dynamic Excel file with data displayed to assess performance in four main areas: key performance indicators, Propensity-Adjusted Interviewer Performance (PAIP) scores, data quality indicators and data set balance.

Key Content Areas of the IPP

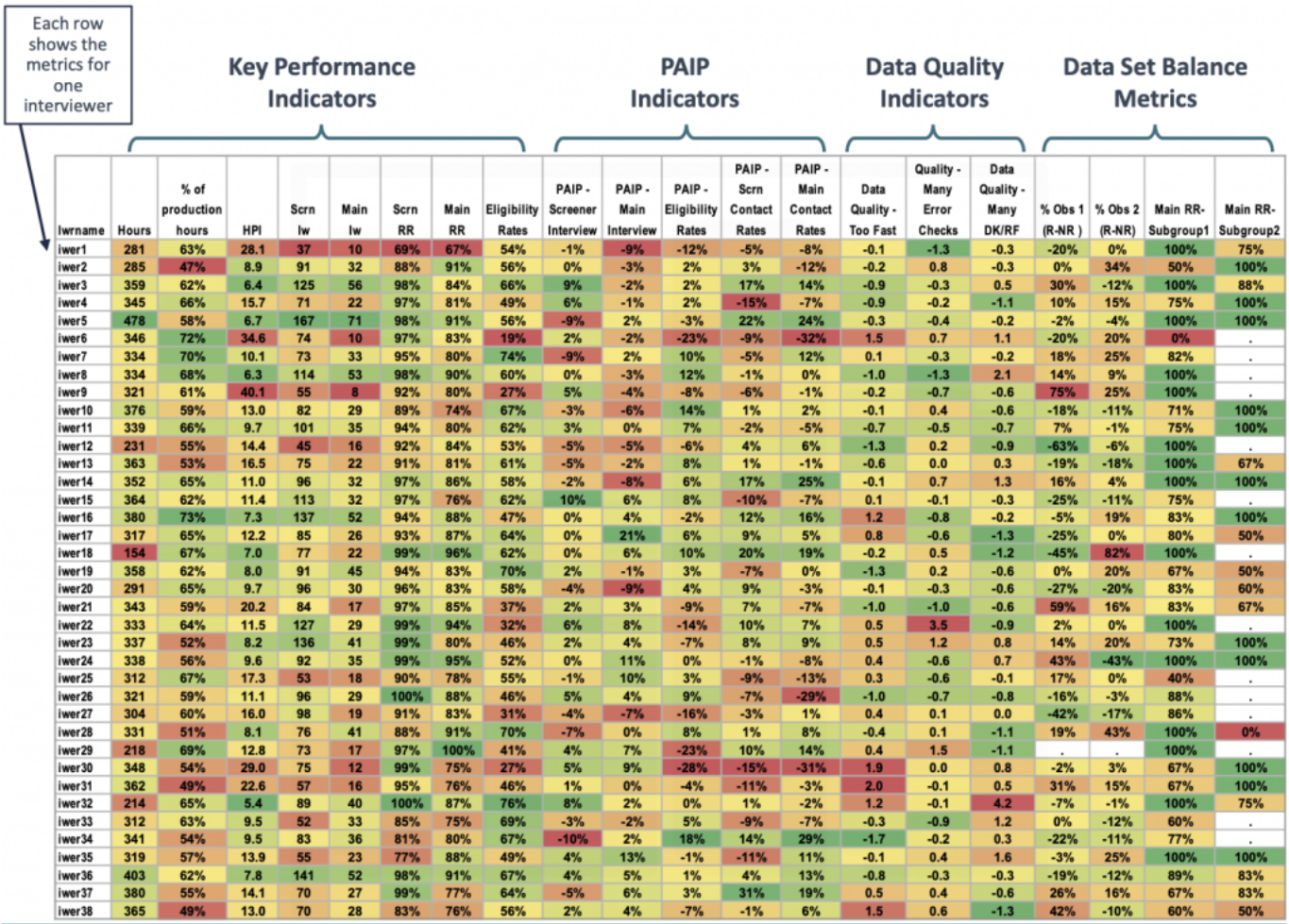
Four areas of performance are assessed in the IPP:

- ***Key Performance Indicators (KPIs)*** evaluate the criteria necessary for the success of the project and provide a means of evaluating individual interviewers as well as the overall project. KPIs included in the IPP to evaluate interviewer effort are: total hours worked to date, hours worked by work type (travel, administrative), hours per completed interview (HPI), interview yield, response rates and eligibility rates. These indicators represent the cumulative performance in the current quarter, as they are calculated based on the cumulative measures from the start date of current quarter to the day before the report was populated. However, the time reference could be adapted to fit the study's objectives (e.g., a full year of data collection).
- ***Propensity-Adjusted Interviewer Performance (PAIP) Scores*** evaluate the effectiveness of the interviewer by incorporating model-based expectations of the *difficulty* of the tasks (West and Groves, 2013). PAIP scores are generated for each interviewer and displayed for interview completion, eligibility and contact.
- ***Data Quality Indicators*** evaluate the quality of the data collected by the interviewer. CAPI was employed by NSFG using Blaise data collection software. Audit trail data captured each key stroke made by the interviewer during each interview, and the time spent between fields. These data was stored and used to create indicators to monitor data quality. Raw indicators include the average time spent on survey questions, the frequency of using help screens, recording remarks, checking errors, backing up during the interview, and the frequency of "don't know" and "refuse" responses. Principle component analysis (PCA) was used to identify common correlations of these measures and reduce the dimensionality of the data, and three factors were identified as data quality indicators: questions being read too quickly, a high proportion of error checks, and a high proportion of refused or don't know responses (Gu et al., 2013).
- ***Data Set Balance Indicators*** provide proxy indicators of non-response bias by examining differences between respondents and non-respondents in terms of key study characteristics.

The data within each performance area is displayed in a heat map format, with gradient coloring (shades of bright green to bright red) to identify positive versus negative performance at a quick glance (Figure 1). Yellow-shaded cells indicate the median value among all interviewers for that specific indicator, whereas green shading indicates positive performance and red shading indicates poor performance. If a larger value of an indicator means better performance (e.g. response rates), then the cell that holds maximum value is highlighted in green and the cells that hold a minimum value are highlighted in red. All other cells are colored proportionally. For indicators in which a smaller value indicates better performance, green and red are coded in reverse while still providing an indicator of poor performance versus good performance. A more extensive review and discussion of each of these components is

described in the supplemental materials and in recent conference proceedings (West, Guyer, Chang AAPOR 2020).

Figure 1. Interviewer Performance Profile (IPP)

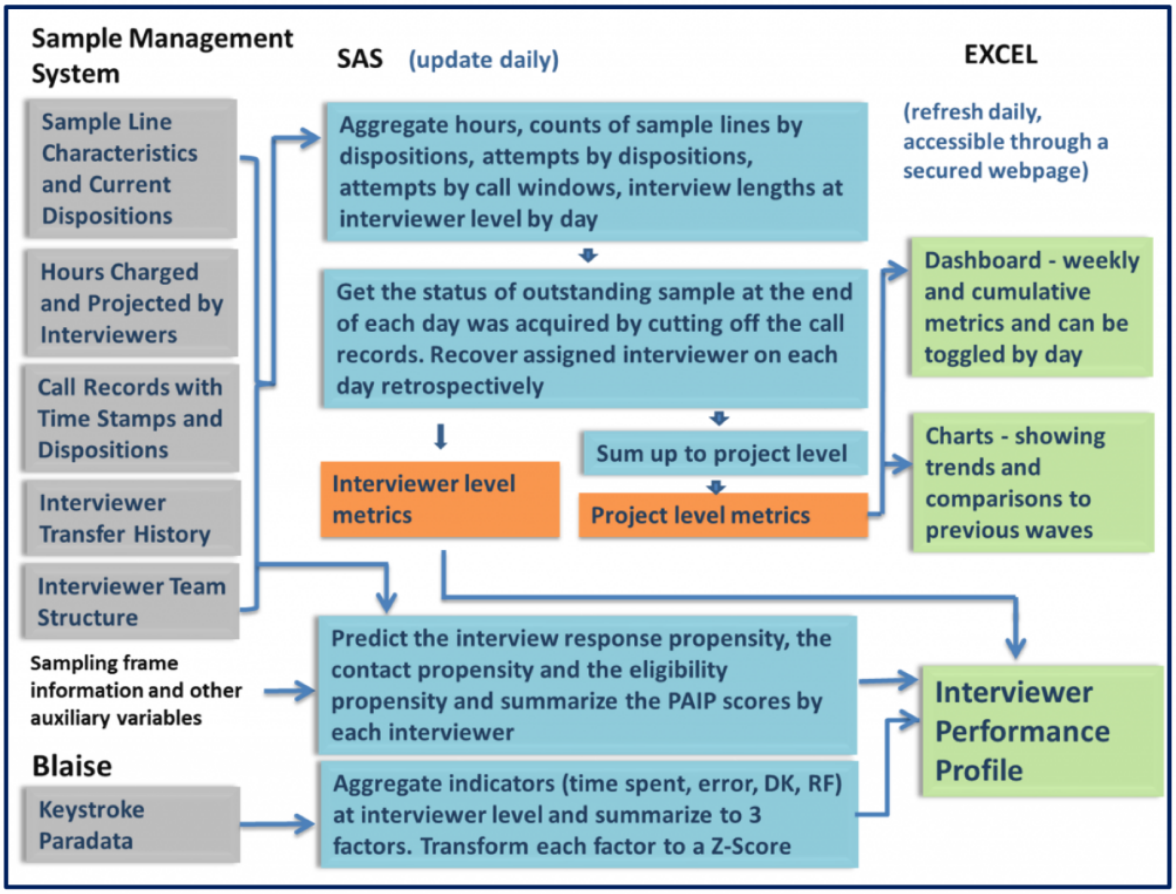


Iwrname	Hours	% of production hours	HPI	Scrn Iw	Main Iw	Scrn RR	Main RR	Eligibility Rates	PAIP - Screener Interview	PAIP - Main Interview	PAIP - Eligibility Rates	PAIP - Scrn Contact Rates	PAIP - Main Contact Rates	Data Quality - Too Fast	Quality - Many Error Checks	Data Quality - Many DK/RF	% Obs 1 (R-NR)	% Obs 2 (R-NR)	Main RR-Subgroup1	Main RR-Subgroup2
iwer1	281	63%	28.1	37	10	69%	67%	54%	-1%	-9%	-12%	-5%	-8%	-0.1	-1.3	-0.3	-20%	0%	100%	75%
iwer2	285	47%	8.9	91	32	88%	91%	56%	0%	-3%	2%	3%	-12%	-0.2	0.8	-0.3	0%	34%	50%	100%
iwer3	359	62%	6.4	125	56	98%	84%	66%	9%	-2%	2%	17%	14%	-0.9	-0.3	0.5	30%	-12%	100%	88%
iwer4	345	66%	15.7	71	22	97%	81%	49%	6%	-1%	2%	-15%	-7%	-0.9	-0.2	-1.1	10%	15%	75%	100%
iwer5	478	58%	6.7	167	71	98%	91%	56%	-9%	2%	-3%	22%	24%	-0.3	-0.4	-0.2	-2%	-4%	100%	100%
iwer6	346	72%	34.6	74	10	97%	83%	19%	2%	-2%	-23%	-9%	-32%	1.5	0.7	1.1	-20%	20%	0%	.
iwer7	334	70%	10.1	73	33	95%	80%	74%	-9%	2%	10%	-5%	12%	0.1	-0.3	-0.2	18%	25%	82%	.
iwer8	334	68%	6.3	114	53	98%	90%	60%	0%	-3%	12%	-1%	0%	-1.0	-1.3	2.1	14%	9%	100%	.
iwer9	321	61%	40.1	55	8	92%	80%	27%	5%	-4%	-8%	-6%	-1%	-0.2	-0.7	-0.6	75%	25%	100%	.
iwer10	376	59%	13.0	82	29	89%	74%	67%	-3%	-6%	14%	1%	2%	-0.1	0.4	-0.6	-18%	-11%	71%	100%
iwer11	339	66%	9.7	101	35	94%	80%	62%	3%	0%	7%	-2%	-5%	-0.7	-0.5	-0.7	7%	-1%	75%	100%
iwer12	231	55%	14.4	45	16	92%	84%	53%	-5%	-5%	-6%	4%	6%	-1.3	0.2	-0.9	-63%	-6%	100%	.
iwer13	363	53%	16.5	75	22	91%	81%	61%	-5%	-2%	8%	1%	-1%	-0.6	0.0	0.3	-19%	-18%	100%	67%
iwer14	352	65%	11.0	96	32	97%	86%	58%	-2%	-8%	6%	17%	25%	-0.1	0.7	1.3	16%	4%	100%	100%
iwer15	364	62%	11.4	113	32	97%	76%	62%	10%	6%	8%	-10%	-7%	0.1	-0.1	-0.3	-25%	-11%	75%	.
iwer16	380	73%	7.3	137	52	94%	88%	47%	0%	4%	-2%	12%	16%	1.2	-0.8	-0.2	-5%	19%	83%	100%
iwer17	317	65%	12.2	85	26	93%	87%	64%	0%	21%	6%	9%	5%	0.8	-0.6	-1.3	-25%	0%	80%	50%
iwer18	154	67%	7.0	77	22	99%	96%	62%	0%	6%	10%	20%	19%	-0.2	0.5	-1.2	-45%	82%	100%	.
iwer19	358	62%	8.0	91	45	94%	83%	70%	2%	-1%	3%	-7%	0%	-1.3	0.2	-0.6	0%	20%	67%	50%
iwer20	291	65%	9.7	96	30	96%	83%	58%	-4%	-9%	4%	9%	-3%	-0.1	-0.3	-0.6	-27%	-20%	83%	60%
iwer21	343	59%	20.2	84	17	97%	85%	37%	2%	3%	-9%	7%	-7%	-1.0	-1.0	-0.6	59%	16%	83%	67%
iwer22	333	64%	11.5	127	29	99%	94%	32%	6%	8%	-14%	10%	7%	0.5	3.5	-0.9	2%	0%	100%	.
iwer23	337	52%	8.2	136	41	99%	80%	46%	2%	4%	-7%	8%	9%	0.5	1.2	0.8	14%	20%	73%	100%
iwer24	338	56%	9.6	92	35	99%	95%	52%	0%	11%	0%	-1%	-8%	0.4	-0.6	0.7	43%	-43%	100%	100%
iwer25	312	67%	17.3	53	18	90%	78%	55%	-1%	10%	3%	-9%	-13%	0.3	-0.6	-0.1	17%	0%	40%	.
iwer26	321	59%	11.1	96	29	100%	88%	46%	5%	4%	9%	-7%	-29%	-1.0	-0.7	-0.8	-16%	-3%	88%	.
iwer27	304	60%	16.0	98	19	91%	83%	31%	-4%	-7%	-16%	-3%	1%	0.4	0.1	0.0	-42%	-17%	86%	.
iwer28	331	51%	8.1	76	41	88%	91%	70%	-7%	0%	8%	1%	8%	-0.4	0.1	-1.1	19%	43%	100%	0%
iwer29	218	69%	12.8	73	17	97%	100%	41%	4%	7%	-23%	10%	14%	0.4	1.5	-1.1	.	.	100%	.
iwer30	348	54%	29.0	75	12	99%	75%	27%	5%	9%	-28%	-15%	-31%	1.9	0.0	0.8	-2%	3%	67%	100%
iwer31	362	49%	22.6	57	16	95%	76%	46%	1%	0%	-4%	-11%	-3%	2.0	-0.1	0.5	31%	15%	67%	100%
iwer32	214	65%	5.4	89	40	100%	87%	76%	8%	2%	0%	1%	-2%	1.2	-0.1	4.2	-7%	-1%	100%	75%
iwer33	312	63%	9.5	52	33	85%	75%	69%	-3%	-2%	5%	-9%	-7%	-0.3	-0.9	1.2	0%	-12%	60%	.
iwer34	341	54%	9.5	83	36	81%	80%	67%	-10%	2%	18%	14%	29%	-1.7	-0.2	0.3	-22%	-11%	77%	.
iwer35	319	57%	13.9	55	23	77%	88%	49%	4%	13%	-1%	-11%	11%	-0.1	0.4	1.6	-3%	25%	100%	100%
iwer36	403	62%	7.8	141	52	98%	91%	67%	4%	5%	1%	4%	13%	-0.8	-0.3	-0.3	-19%	-12%	89%	83%
iwer37	380	55%	14.1	70	27	99%	77%	64%	-5%	6%	3%	31%	19%	0.5	0.4	-0.6	26%	16%	67%	83%
iwer38	365	49%	13.0	70	28	83%	76%	56%	2%	4%	-7%	-1%	6%	1.5	0.6	-1.3	42%	-10%	60%	50%

Statistical Processing Steps

The statistical processing steps are shown in Figure 2. Data was extracted from several data tables derived from the sample management system including a sample table, an effort table, and a table with sample outcomes as measures of productivity. Once each of the data sets are compiled, they were then combined into a single file and the daily and cumulative metrics are calculated for each interviewer as described above. For comparison purposes, the data are also summed at the project level to determine the proportion or average of each indicator at the project level by day. The variables were then placed in the order needed for the charts to evaluate the four performance areas. The statistical software package SAS v9.4 (Cary, NC) was used to process the data utilized in the IPP and to export the IPP data to Excel.

Figure 2. IPP Data Flow



IPP Case Studies

The IPP can be used to evaluate performance at a single point in time as well as to track improvements in performance after interventions occur. Interventions may include covering topics on team calls or in emails to the full study team if performance issues seem widespread, such as low response rates with specific subgroups of cases, high hours per complete for the quarter, or increased resistance rates. We now provide examples of performance indicators that led to concerns, the interventions that occurred, and the performance following the intervention as compared to the initial performance indicators observed.

Table 1 displays results for six interviewers, A-F, at three different points in time as indicated in the second column “Week” (week of the quarter). The total hours worked through that week are displayed as well as the number of completed screener and main interviews, the cumulative hours per interview (HPI) at that point in time, the eligibility rate, the screener response rate (Screen RR) and the main interview response rate (Main RR). Areas of concern for each interviewer are highlighted in red, whereas areas of positive performance are highlighted in green. It should be noted that the data displayed in the table are fictitious in order to preserve interviewer anonymity. The case studies illustrate similar scenarios that have occurred while using the IPP yet merely serve as examples of the various ways in which the IPP can be used.

Table 1. Case Study Examples

Interviewer	Week	Total Hours	# Screener Interviews	# Main Interviews	HPI	Eligibility Rate	Screener RR	Main RR	Notes
A	3	93	25	2	46.5	34	22	23	New hire; performance discussion
A	5	152	50	3	50.7	34	52	18	2nd performance discussion
A	10	289	82	8	36.1	35	76	28	Some improvement noted
B	3	4	0	0	0	0	0	0	New hire; performance discussion
B	5	44	30	6	7.3	70	22	29	2nd performance discussion
B	10	176	53	25	7.0	66	51	41	Improvement noted
C	3	39	5	2	19.5	90	5	60	New hire; performance discussion
C	5	67	15	3	22.3	90	15	20	2nd performance discussion
C	7	125	25	8	15.6	86	75	60	Resignation submitted
D	3	68	12	1	68.0	72	18	10	New hire; performance discussion
D	5	118	22	2	59.0	70	38	12	2nd performance discussion
D	10	200	26	3	66.7	70	62	18	Termination
E	3	144	60	20	7.2	43	60	75	Experienced iwer
E	5	182	70	24	7.6	45	75	72	Performance maintained
E	10	370	118	52	7.4	50	95	82	Optimal performance noted
F	3	105	40	20	5.3	85	100	100	Experienced traveller
F	5	110	40	20	5.5	85	100	100	No trips this week
F	10	225	80	35	6.4	85	99	100	Optimal performance noted

HPI: Hours per Interview; RR: Response rate

Interviewer A

This was a new interviewer who started on the NSFG for the first time in this quarter. By the third week of the quarter, the interviewer had worked 93 hours. NSFG interviewers are required to work 30 hours per week, so the level of effort, in terms of hours charged, was appropriate. The number of completed screening interviews was 25. At a 34% eligibility rate, nine eligible main respondents would have been identified. However, main interviews had only been completed with two of the nine, thus her HPI was extremely high (46.5 compared to project average of 10.6). A performance discussion with this interviewer took place at the end of week 3. No improvement was observed two weeks later in week 5. This interviewer appeared to be focusing on screening rather than conducting main interviews. A second performance discussion took place in week 5. By week 10, some improvement was observed: six additional main interviews had been completed, thus lowering her HPI, and she continued to focus on interviewing. This interviewer showed some improvement with additional coaching in the first quarter and was retained on the project. Her supervisor continued to monitor her performance.

Interviewer B

This was also a new interviewer who started on the NSFG this quarter. By week three, the interviewer had only worked four hours and had not conducted any field work. A performance discussion took place with this interviewer during week 3. By week 5, she had started working, although still not at the level required. However, her effort and productivity was on track and her HPI was quite good. Despite these outcomes, a performance discussion was held with the interviewer in week 5 as she was not yet working the required number of hours. This interviewer's performance had improved significantly in all areas by week 10. In fact, she showed strong potential in completing main interviews and maintaining a low HPI. With some additional coaching in the first quarter, this new interviewer is now well on her way to a successful new career.

Interviewer C

This was a new interviewer who started on the NSFG this quarter as well. A performance discussion was held in week 3 as she was not meeting the requirement of working thirty hours per week and her production outcomes were low for this point in time. By week 5, her productivity had declined further. She

was not working the required number of hours and was thus unable to achieve an acceptable level of productivity. A second performance discussion took place in week 5. This interviewer resigned in week 7.

Interviewer D
This was also a new interviewer who had performance discussions with her supervisor in weeks 3 and 5 due to low performance and working an insufficient number of hours. Her performance did not improve over the ten week period despite receiving additional coaching and performance improvement notifications. This interviewer was terminated from the project in week 10. Although she did complete some work, she would be a risk to the project were she to continue and would incur increased project costs related to monitoring and coaching without showing the improvements (no/low return on investment).

Interviewer E
This experienced interviewer was a high-performing interviewer who had worked on the project for more than five years prior to the current year of data collection. She had consistently high performance in completing the screener and main interviews, maintained her hourly commitment, and consistently had a lower than average HPI. This interviewer did not require additional coaching or performance discussions, nor did she introduce increased cost to the project. Instead, she was able to provide guidance to new, less experienced interviewers and her efficiency led to decreased project costs.

Interviewer F
This experienced interviewer had inconsistent performance in the first five weeks of the quarter as she did not have any sample in her area. Instead, she travelled to other areas to conduct interviews in situations where the local interviewer was on leave or was unable to work their full sample load, as well as to unstaffed areas (e.g., due to interviewer attrition) and always completed her assigned sample while traveling. However, only interviewers with historically good performance indicators are eligible to work as “travelers”. The primary goal of the traveler is to be as efficient as possible in the weeks that they are traveling. This interviewer maintained high levels of productivity and a low HPI and maintained her status as a high performer by the end of the quarter.

Conclusion/Discussion

As these examples show, the IPP is a useful tool for both the interviewer as well as the project team. The IPP is a management tool, enabling the evaluation of numerous performance metrics across multiple interviewers simultaneously. It allows managers to identify areas of concern, intervene, and monitor outcomes, contributing to the overall success of the project. The IPP allows interviewers who are struggling to receive additional coaching and to improve their performance, thus providing employment stability. Additionally, the IPP highlights interviewers with exceptional performance and allows those interviewers to be recognized and to contribute to the coaching of new interviewers to ensure their long-term success as well.

The dynamic nature of the IPP has allowed for the integration of new measures over time. While there is an initial investment at the start of a project in defining the elements to monitor and the paradata needed to develop the IPP, this investment can lead to cost reductions in key areas of production over the course of the study. For example, identifying and intervening with interviewers with low response rates or high HPIs can lead to reduced costs and improved production parameters, if detected early enough and if the interventions are successful. The IPP can aid in early detection and tracking of intervention outcomes. Additionally, once the tool is developed, it can be further refined and adapted for other studies as well. As

in previous reports, the use of heat maps allows for the identification of positive versus negative performance at a quick glance. The reviewer can then dig deeper to determine the cause of the poor performance and develop an action plan. We have found the dynamic IPP to be incredibly useful in monitoring interviewer performance, identifying areas for improvement, and tracking changes over time.

The IPP is a useful tool for studies that are smaller or more narrow in size and scope—such as the number of interviewers, cases, or parameters measured– as well as those that are larger in size or scope. A minimum number of cases is not required in order for the tool to be useful. Instead, the key is providing the sample sizes for which the indicators are based to allow the managers to make decisions as to whether to put weight on the values of the indicators. The indicators are meant to be reviewed together with other indicators on the report. Instead of voiding the indicators due to a small number of cases, we suggest reporting the denominators of the rates that tend to have a smaller sample size on the report and allowing managers to use the information accordingly. The indicators are color-coded based on the standard deviation of the mean of all interviewers to help identify the outliers of each indicator. Thus, the IPP helps to provide a quick glance at all perspectives of each interviewer compared to other interviewers. This approach can be used for larger studies as well where the performance of all interviewers can be assessed in comparison to the full team, or subgroups of interviewers or cases can be assessed separately. For example, the performance of a regional teams of interviewers could be compared to the other interviewers within the team to determine those who are excelling in that specific geographic region and those who may need additional training or guidance. Additionally, subsets of cases could be evaluated, such as by sample release or other indicators of key sample characteristics such as cohort, age, or sample source.

Our work as survey researchers must continue to adapt to both changes in technology and provide new views of the available data to ensure efficient data collection and successful outcomes. The utility of paradata to measure and monitor these trends, and the visual display of data, will continue to expand over time. The IPP is an easy tool for both methodologists and managers to use to increase productivity and efficiency while also decreasing costs, measurement error and non-response error.

Appendix

Supplemental material

References

1. Biemer, P., Herget, D., Morton, J., and Willis, G. (2000). The feasibility of monitoring field interview performance using computer audio recorded interviewing (CARI). Proceedings of the American Statistical Association’s Section on Survey Research Methods, pp. 1068-1073.
2. Durand, C. (2005). Measuring Interviewer Performance in Telephone Surveys. Quality and Quantity. 39. 763-778. 10.1007/s11135-004-6781-0.
3. Durand, C. (2008). Assessing the Usefulness of a New Measure of Interviewer Performance in Telephone Surveys. Public Opinion Quarterly. 72(4)741-752.
4. Groves, R.M. and Heeringa, S.G. (2006). Responsive design for household surveys: tools for actively controlling survey errors and costs. Journal of the Royal Statistical Society: Series A (Statistics in Society), Vol. 169 No. 3, pp. 439-457.
5. Groves, R.M., and McGonagle, K. (2001). A Theory-Guided Interviewer Training Protocol Regarding Survey Participation. *Journal of Official Statistics,* 17(2): 249-266.
6. Gu H., Couper M., Kirgis N., Parker S., and Buageila S. (2013). Using Audit Trail Data for Interviewer

Data Quality Management. AAPOR 2013 Annual Meeting Proceedings.
7. Guyer, H. (2004). How can we predict and plan for successful contact attempts? International Field Directors and Technologies Conference (IFDTC 2004). Scottsdale, AZ.
8. Kroke A., Bergmann M.M., Lotze G., Jeckel A., Klipstein-Grobusch K., Boeing H. (1999). Measures of Quality control in the German component of the EPIC study. Annals of Nutrition and Metabolism 43:216–224.
9. Laflamme, F. and St. Jean, H. (2011). Proposed indicators to assess interviewer performance in CATI surveys. Proceedings of the Survey Research Methods of the American Statistical Association, Joint Statistical Meetings, Miami, Florida, August 2011.
10. Mockovak, W., Yu, A., Earp, M. (2015). Using Calibration Training to Assess the Quality of Interviewer Performance. Bureau of Labor Statistics (BLS).
11. Mohadjer, L. and Edwards, B. (2018). Paradata and dashboards in PIAAC. Quality Assurance in Education, 26(2), 263-277.
12. Olson, K. M., & Bilgen, I. (2011). The role of interviewer experience on acquiescence. Public Opinion Quarterly, 75(1), 99-114.
13. Powell, M.B., Cavezza, C., Hughes-Scholes, C., and Stoove, M. (2010). Examination of the consistency of interviewer performance across three distinct interview contexts. Psychology, Crime & Law, 16:7, 585-600
14. Sudman, S. (1967). Quantifying interviewer quality. Public Opinion Quarterly. 30(4), 1967. 664-667.
15. Wagner, J., & Guyer, H. (2005). Statistical propensity models to predict likelihood of survey response compared to interviewer judgments of likelihood of response. Midwest Association for Public Opinion Research. Chicago, IL.
16. Wagner, J., West, B. T., Kirgis, N., Lepkowski, J. M., Axinn, W. G., Ndiaye, S. (2012). Use of paradata in a responsive design framework to manage a field data collection. Journal of Official Statistics 28:477–99.
17. West, B.T., Groves, R.M. (2013). A Propensity-Adjusted Interviewer Performance Indicator. Public Opinion Quarterly. 77(1), 2013. 352-374.
18. West, B. T., Kreuter, F. A practical technique for improving the accuracy of interviewer observations: Evidence from the National Survey of Family Growth. 2015. Field Methods 27(2) 144-162.
19. West, B. T., Guyer, H.M., Chang, W. Using Paradata to Develop and Implement an Interviewer Performance Profile for Monitoring and Evaluating Interviewer Performance. AAPOR 2020 (Virtual Conference Presentation).