Describing Candidate Assessments and Hiring Outcomes at DOI

An evaluation summary for the Department of the Interior

Office of Evaluation Sciences
US Department of the Interior
January 2023

Table of Contents

Introduction

Executive Order (EO) 13932 "Modernizing and Reforming the Assessment and Hiring of Federal Job Candidates" was designed to change the application and screening process which has, historically, relied heavily on applicant self-assessments of technical competencies. EO 13932 requires agencies to use alternative methods to assess applicants' qualifications for federal jobs; essentially, applicants must clear alternative assessment hurdles in order to be considered qualified and eligible for preference and referral.

The Department of Interior (DOI) began implementing EO 13932 in 2021 by offering two available assessment methods to be applied to hiring actions: (1) using subject-matter experts (SMEs) to manually evaluate candidate resumes; or (2) using USA Hire scored assessments of general competencies. These assessment methods were adopted and coupled with questionnaires typically used for candidates to self-assess their technical competencies.

The objective of this descriptive evaluation was to explore hiring outcomes across candidate assessment methods over the past two years.

There are three primary research questions:

- How are candidate assessments being implemented at DOI?
- How long does the hiring process take and how often are positions filled for each candidate assessment method?
- Does the representation of female, people of color, and Veteran candidates change at key stages of the recruitment process within each assessment method?

To answer these questions, we employed quantitative statistical analysis of administrative data to explore means and proportions of outcome variables of interest. The evaluation includes analyses of hiring actions and applications from 10 DOI critical occupations, or job series,[1] within five DOI bureau/offices.[2] Vacancies included in the data are permanent, non-supervisory delegated examining (Public Non Status) positions announced between July 2020 and July 2022.

The evaluation is descriptive and not intended to determine the relative effectiveness of different candidate screening mechanisms; instead, descriptive results provide a baseline understanding of hiring outcomes for hiring actions that use each type of assessment. Although the findings do not support any causal inferences about the effects of assessment methods on recruitment outcomes, they help describe the implementation of the three candidate assessment tools at DOI.

[1] Occupational series included: 0401: General Natural Resources Management and Biological Sciences Series; 0025: Park Ranger; 0462: Forestry Technicians; 1315: Hydrology; 0810: Civil Engineering; 1801: General Inspection, Investigation, Enforcement, and Compliance Series; 1350: Geology series; 1811: Criminal Investigations ; 0028: Environmental Protection Specialist; 0802: Engineering Technical.
[2] Bureaus included: U.S. Geological Survey (USGS), National Park Service (NPS), Bureau of Land Management (BLM), Forestry & Wildlife Services (FWS), Bureau of Reclamation (BOR).

Candidate Assessment Tools

As a vacancy announcement is being prepared, hiring officials work with human capital specialists to identify core job competencies, develop the position description, and choose a method for assessing applicants. Assessments are chosen based on the job requirements, estimated size of the applicant pool, and the urgency of needing a position filled.

We explore three different candidate assessment tools:

Self-assessments of technical competencies: Self-assessments allow applicants to self-report their experience, expertise, and competencies related to the job announcement on a questionnaire. Self-reported ratings are then reviewed by human capital specialists to ensure that information in the application package supports the self-reported ratings.

Subject-matter expert (SME) panel manual assessments: Manual assessments are administered outside of the staffing system and inputted into the hiring system manually, based on the expert review scores. Manual assessments may include writing tasks, structured interviews, rating schedules, or structured Subject Matter Expert (SME) resume reviews. DOI convenes subject-matter experts to review applications and score applicants based on evidence of qualifications for the position.

USA Hire scored assessments: USA Hire is the Office of Personnel Management (OPM) run assessment and evaluation service that is a part of the USA Staffing platform, the federal government's integrated talent acquisition system. USA Hire offers 118 assessments of general competencies which can be scored to determine applicants' eligibility for a specific job. Assessments are tied to particular job series and grade levels. These assessments measure things like reasoning skills, decision-making ability, math skills, interpersonal skills, stress tolerance, etc. USA Hire aims to provide a "whole person" assessment to inform hiring.

Priority Hiring Outcomes

Priority hiring outcomes for DOI include the duration of time for key milestones in the recruitment process (i.e., creating a list of qualified and eligible candidates, selecting a candidate from the eligible list), the rate of hiring success, and the proportion of applicants with different characteristics at each stage of the recruitment process.

We describe six different outcomes:

Certification time: Duration of time (in days) between the closing of the vacancy application period and the date that a certificate of eligible applicants (the "cert list") is created.

Post-certification time to selection: Duration of time (in days) between the issuance of the cert list and the first offer made to a candidate (for successful searches) or cancellation of the search (for unsuccessful searches).

Recruitment success: Binary indicator of whether or not an offer was made to a candidate from the cert list. Success indicates at least one offer was made; failure indicates no offer was made or the search was canceled.

Female applicants: Proportion of applicants that identify as female among completed applications, certified eligible list, and those selected for the position.

Applicants of color: Proportion of applicants who identify as non-White or Hispanic among completed applications, certified eligible list, and those selected for the position.

Veteran applicants: Proportion of applicants adjudicated with Veteran's preference among completed applications, certified eligible list, and those selected for the position.

Figure 1. Process Map of Key Stages in the Hiring Process

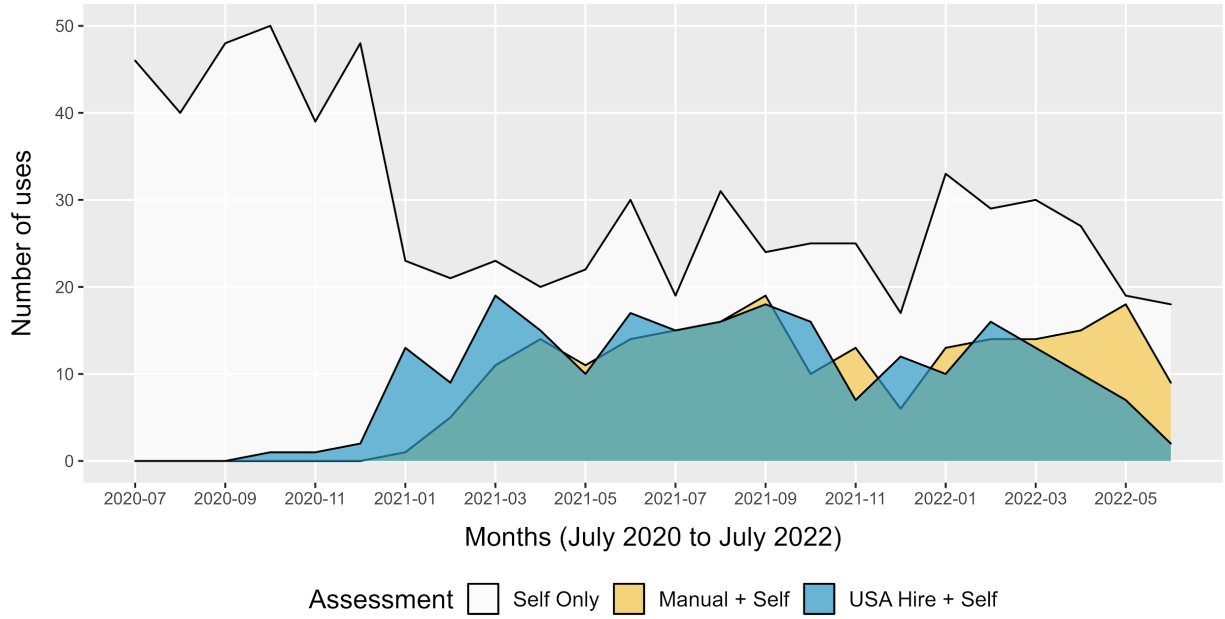
Summary of Results

Research Question 1: Trends in the implementation of candidate assessment tools

To describe the trends in candidate assessments implementation across DOI, we illustrated utilization rates for each of the three assessment types using ridgeline plots. Three analyses examined 1) the rate of use for all three assessment types across DOI, 2) the rate of use for all three assessment types for select bureaus in DOI, and, 3) the rate of use for all three assessment types across select DOI job series, each by month.

Results show that prior to January 2021 DOI primarily relied on self-assessments of technical competencies; after January 2021 self-assessments decreased in frequency as the department adopted manual and USA Hire assessments (Figure 2) (N = 1,154). Further description of assessment tools by bureau and job series can be found in Appendix A.

Figure 2. Implementation of Hiring Assessments Across DOI

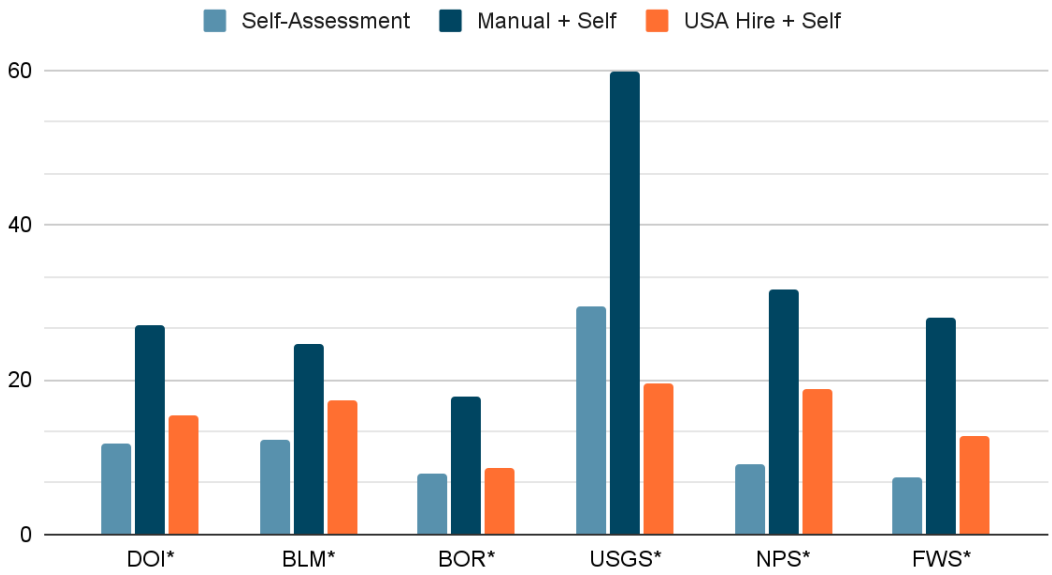


Research Question 2: Timing and success of recruitment process

For the second research question we observed the length of time between key hiring milestones for DOI hiring actions. We examined the average certification time, post-certification time to selection, and recruitment success for each assessment type. Overall, the average time to certify candidates was 15.3 days and the time between certification and selection was 37.1 days. Across DOI 69% of hiring actions resulted in a selection.

The time from the close of vacancy announcements to the cert list being issued averages about 12 days for self-assessments, 27 days for manual assessments, and 15 days for USA Hire assessments. Across bureaus average certification time ranged from 7 days (FWS using self-report only assessments) to 60 days (USGS using manual assessments). Although average hiring outcomes may differ across assessment types, these differences may be a result of various components of the hiring process, and cannot be attributed to the assessment methods themselves.

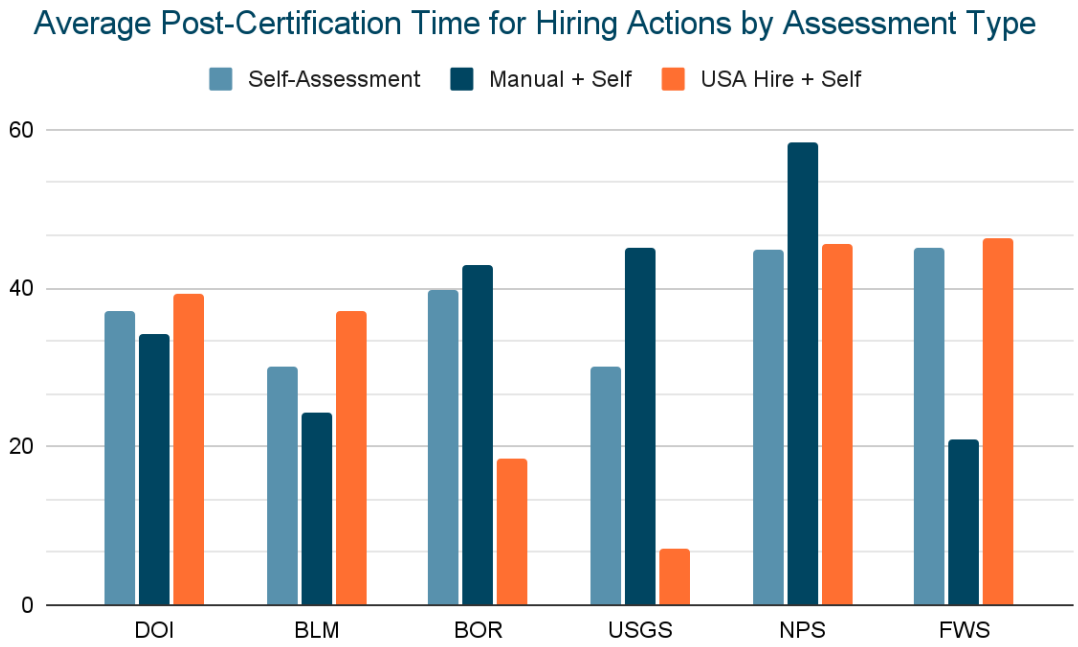
Figure 3. Hiring action certification time for DOI and bureaus, by assessment type



Note: As an exploratory analysis we conducted a one-way analysis of variance (ANOVA) for DOI and each bureau. An asterisk (*) indicates that an F-test rejects the null hypothesis that certification time is the same across assessment methods (at p-value <0.05). Details on the exploratory analysis are described below.

The majority of the recruitment time is accounted for in the post-certification period for all assessment methods. The average time from the cert list being issued to selection is 37 days for self-assessments, 34 days for manual assessments, and 39 days for USA Hire assessments. Across bureaus post-certification time ranged from 7 days (USGS using USA Hire assessments, although this only included six hiring actions) to 58 days (NPS using manual assessments).

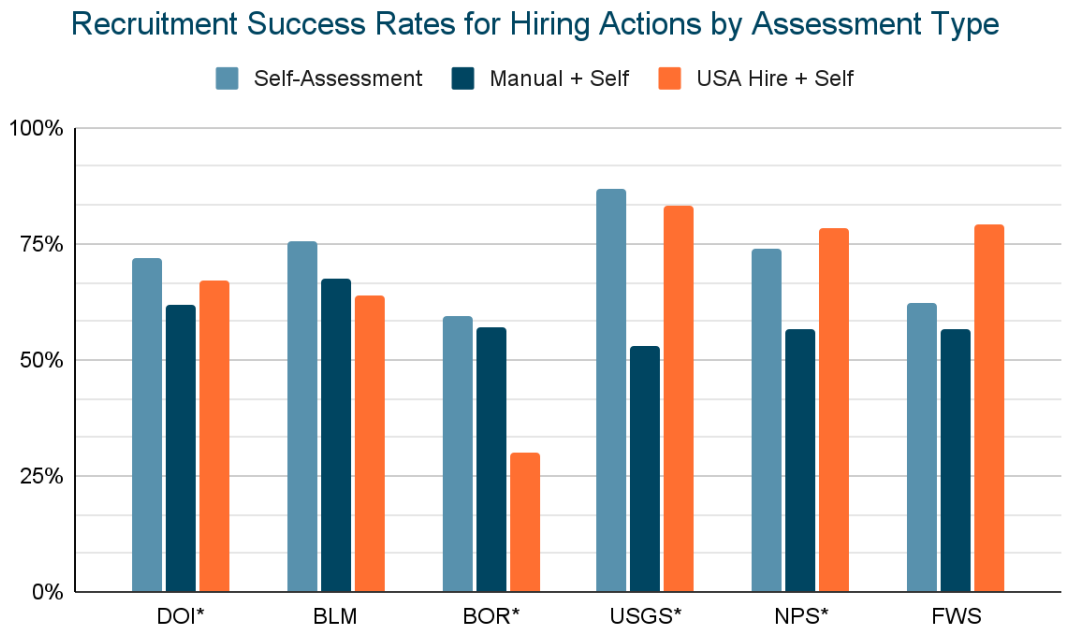
Figure 4. Hiring action post-certification time for DOI and bureaus, by assessment type



Note: As an exploratory analysis we conducted a one-way analysis of variance (ANOVA) for DOI and each bureau. An asterisk (*) indicates that an F-test rejects the null hypothesis that post-certification time is the same across assessment methods (at p-value <0.05). Details on the exploratory analysis are described below.

Overall, recruitment success for DOI ranges between 60% and 70%. Self-report assessments have the highest recruitment rates with 72% of hiring actions resulting in a selection, followed by hiring actions using USA Hire (67%) and those using manual assessments (62%). The highest recruitment success rate was 87% (USGS using self-report only assessments) and the lowest recruitment success rate was 30% (BOR using USA Hire assessments).

Figure 5. Recruitment success rate for DOI and bureaus, by assessment type



Note: As an exploratory analysis we conducted a one-way analysis of variance (ANOVA) for DOI and each bureau. An asterisk (*) indicates that an F-test rejects the null hypothesis that recruitment success is the same across assessment methods (at p-value <0.05). Details on the exploratory analysis are described below.

Exploratory analysis of differences in hiring action-level outcomes by assessment method

To explore whether there are differences in certification time, post-certification time, and recruitment success across assessment types we conduct three one-way analysis of variance (ANOVA) tests for hiring actions within DOI. Differences that exist between assessment methods do not explain *why* differences may exist. For example, the differences may be a result of various components within the recruitment process, such as hiring manager choices, and can not be directly attributed to the assessments used and are not measured in this evaluation.

The ANOVAs indicate that there are statistically significant differences in certification time ($F=55.3$, $df= 2$, $p = <0.001$) and recruitment success ($F=4.08$, $df= 2$, $p = 0.02$) across assessment methods (Table 1). Post-certification time does not differ significantly across assessment methods ($F=0.43$, $df= 2$, $p = 0.648$).[3] To better understand the significant differences, aTukey's post-hoc

[3] We also conducted ANOVAs on the hiring action-level outcomes for each of the five bureaus. The ANOVAs reject the hypothesis that mean outcomes are the same for certification time only for the Bureau of Land Management and the Fish and Wildlife Service. The null hypothesis is rejected for certification time and recruitment success for the Bureau of Reclamation and the U.S. Geological Survey. Full results are presented in Appendix B.

analysis was used to compare pairwise differences in mean certification time and recruitment success for the three assessment methods.[4]

For certification time all three of the pairwise differences between assessment methods were statistically significant. Average certification time was 15.2 days shorter for self-assessments compared with manual assessments (p = <0.001), 11.7 days shorter for USA Hire assessments compared with manual assessments (p = <0.001), and 3.5 days longer for USA Hire assessments compared with self-assessments (p = .03).

Recruitment success was 9.9 percentage points higher for self-assessments compared with manual assessments, and this difference was statistically significant (p = .02). Differences in recruitment success were not statistically significant for USA Hire compared with manual assessments or USA Hire compared with self-assessments.

Table 1. One-way ANOVA for DOI hiring action-level outcomes

Outcome	***F statistic (df = 2)***	***p-value***	**Tukey's post-hoc analysis of pair-wise mean differences in outcomes**		
			Self vs. Manual	*USA Hire vs. Manual*	*USA Hire vs. Self-Report*
Certification time (days)	55.31	<0.001	-15.2 (*p* <0.001)	-11.7 (*p*<0.001)	3.51 (*p* = 0.031)
Post-certification time (days)	0.43	0.648	*n/a*	*n/a*	*n/a*
Recruitment success (%)	4.08	0.017	0.099 (*p* = 0.015)	0.053 (*p* = 0.442)	-0.046 (*p* = 0.388)

Note: We reject the hypothesis of no differences in the outcome across assessment methods if the p-value from the F-test is less than 0.05.

Research Question 3: Representation of candidates in key stages of the recruitment process

We calculated the proportion of applicants at each stage of the recruitment process who are female, non-White or Hispanic, and are Veterans with preference. These proportions were calculated for the entire pool of applicants, those who made the cert list, and those who were selected. Sex, race, and ethnicity are characteristics that applicants self-report (but can opt out of reporting); we calculated proportions among applicants that chose to report demographic characteristics and veteran status.[5] All applicants are designated as either Veterans with

[4] We pre-specified in the analysis plan that we would compare pairwise mean outcomes for any outcome where the F-test rejected the null hypothesis that differences were not significant, but we would not compare pairwise means where the null hypothesis was not rejected.

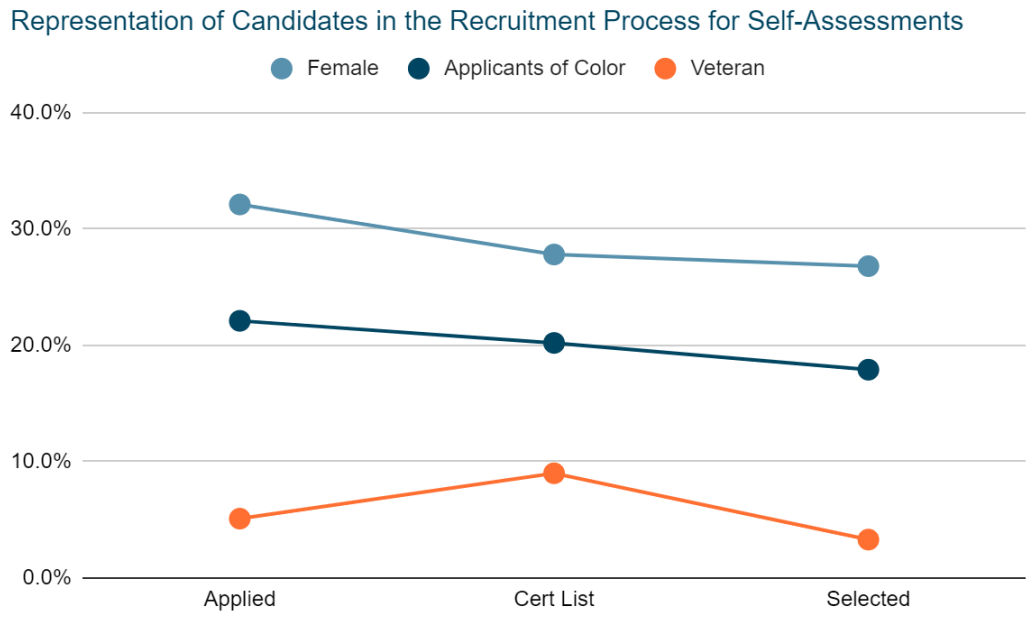
[5] Between 30% and 33% of applicants choose not to report demographic characteristics. Applicants who do not report demographic characteristics do not appear to be more or less likely to be included in subsequent stages of the recruitment process.

preference or no Veteran preference.[6] For each applicant characteristic (female, non-White or Hispanic, Veterans with preference) we conducted a chi-square goodness-of-fit test of whether the proportion of applicants at each stage is statistically different from the other stages.

There were 115,727 applications across the five largest bureaus at DOI (n_{BLM} = 36,560, n_{BOR} = 4,754, n_{GS} = 15,448, n_{NPS} = 23,590, n_{FWS} = 35,375).[7] Of the 80,191 applicants who chose to report their sex, 26,782 (33.4%) were female applicants. Of the 79,242 applicants who reported their race or ethnicity, 17,349 (21.9%) were from applicants of Color. A total of 6,242 (5.4%) applicants received Veterans' preference.

For self assessments, the proportion of applicants who are women decreases at each stage of the recruitment process, as does the proportion who are non-White or Hispanic. For Veterans with preference, representation increases at the cert list stage, but decreases at the selection stage. For all three groups the proportion among selected is less than at the application stage, and these differences are statistically significant (Table 2).

Figure 6. Proportion of applicants at each stage of the recruitment process for self assessments, by candidate characteristics

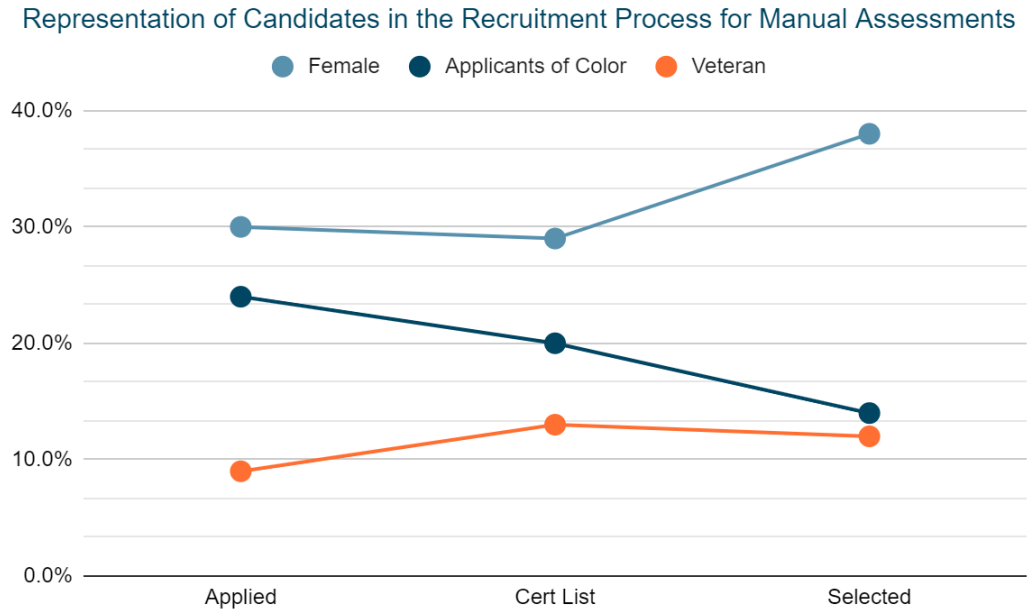


For hiring actions that use manual assessments, the proportion of women among those on the cert list is almost unchanged from the proportion of those who applied, however the proportion of women among the selected is about 8 percentage points higher. This difference is statistically

[6] Applicants can claim preference as a Veteran in the application stage and provide documentation of Veteran status. This claim is then adjudicated by human capital specialists. We tabulate the proportion of Veterans in the recruitment process based on adjudicated status; some applicants who claim status do not receive preference adjudication.
[7] As part of an exploratory analysis we also tabulated applicant proportions by assessment method for the five DOI bureaus included in the study. These tabulations are presented in the appendix. We do not conduct statistical tests of differences in proportions for the bureau-level tabulations.

significant. The proportion of Veterans who make the cert list and are selected increases slightly compared to those who applied, but these increases are not statistically significant. Finally, the proportion of non-White or Hispanic applicants significantly decreases at each stage, by a total of 10 percentage points from application to selection.

Figure 7. Proportion of applicants at each stage of the recruitment process for manual assessments, by candidate characteristics



For hiring actions that use USA Hire assessments, the representation of women slightly increases at each stage, although the difference in the proportion of women among selected is not a significant change. Veterans representation significantly increases at the cert list stage compared to applications, then significantly decreases at the selection stage. The representation of non-White or Hispanic applicants significantly decreases at the cert list stage, and is slightly lower among those selected than at the application stage. This decrease was not significant.

Figure 8. Proportion of applicants at each stage of the recruitment process for USA Hire assessments, by candidate characteristics.

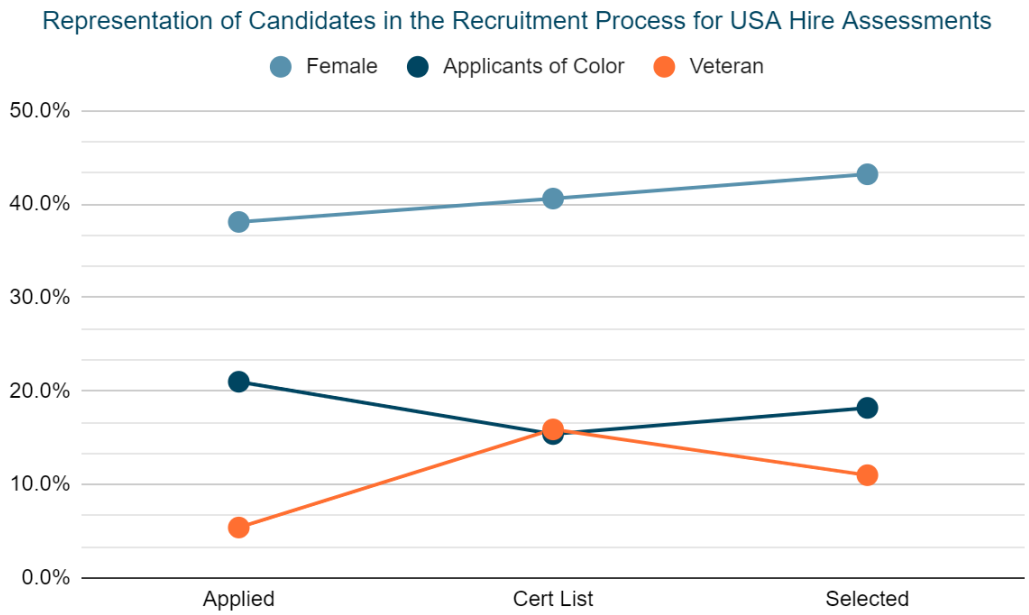


Table 2. Differences in the proportion of candidates at each stage of the recruitment process, by applicant characteristics and assessment method

Self-reported assessments			
	Difference in proportion between stages (χ^2)		
Candidate characteristic	*Applied to cert list*	*Cert list to selected*	*Applied to selected*
Female	-0.04* (196.43)	-0.01 (1.22)	-0.05* (30.67)
Non-White or Hispanic	-0.01* (46.00)	-0.02* (7.45)	-0.04* (23.06)
Veteran	0.04* (1,048.62)	-0.06* (136.80)	-0.02* (22.71)
Manual assessments			
	Difference in proportion between stages (χ^2)		
	Cert-list - Applied	*Selected - Cert-list*	*Selected - Applied*
Female	-0.018 (0.36)	0.09* (6.90)	0.08* (5.70)

Non-White or Hispanic	-0.04* (9.30)	-0.06* (3.87)	-0.10* (8.82)
Veteran	0.04* (40.65)	-0.01 (0.42)	0.03 (2.45)
USA Hire assessments			
	Difference in proportion between stages (χ^2)		
	Cert-list - Applied	*Selected - Cert-list*	*Selected - Applied*
Female	0.03* (10.58)	0.03 (0.92)	0.05 (3.60)
Non-White or Hispanic	-0.06* (72.03)	0.028 (1.84)	-0.03 (1.42)
Veteran	0.11* (1,111.34)	-0.05* (7.85)	0.06* (27.08)

Note: * indicates that a p-value is low enough to classify a result as statistically significant after adjusting p-values to hold the family-wise error rate (FWER) at 5% across all three tests for each assessment-characteristic combination. We adjusted p-values using a Holm-Bonferroni procedure.

Limitations

This descriptive analysis is limited to the scope and data included in the study. The scope of the study included DOI bureaus that have the greatest number of hiring actions and the job series that have been identified as mission critical for the agency. Although an informed design decision, this study's findings may not be representative of other bureaus and job series at DOI.

The administrative data used for the study presented some limitations for the analysis. First, merging the datasets resulted in some unmatched records of hiring actions. It is possible that there are additional hiring actions that could be included in the study if data fields for the appropriate outcomes were matched. Second, some hiring actions did not indicate recorded decisions. For these hiring actions it is unknown whether the recruitment effort has been canceled without a selection or is still open for selection.

Finally, tabulations of applicant demographic characteristics (sex, race, and ethnicity) rely on self-reporting by applicants. About 70% of applicants in this study submitted demographic information. It is possible that unobserved variation in demographic characteristics among the 30% of applicants who don't self-report could change the estimated distributions of candidate characteristics at each stage in the recruitment process.

Conclusion

The results of the descriptive study provide a baseline understanding of DOI's implementation of EO-compliant assessments and how key hiring outcomes vary across the population of hiring

actions. Although the findings do not support any causal inferences about the effects of assessment methods on recruitment outcomes, they help describe the implementation of the three candidate assessment tools at DOI.

DOI began widespread adoption of manual and USA Hire assessments in January 2021. The evaluation uses data on hiring actions from July 2020 to July 2022 to summarize the time it takes to identify a list of qualified candidates and select a candidate and the likelihood of a successful search. These measures vary widely across assessment methods, bureaus, and occupations. The evaluation also examines application data to describe the representation of women, non-White or Hispanic, and Veterans at each stage of the recruitment process.

Future exploration in this area could examine reasons self-reported assessments are still used relatively frequently or why manual assessments or USA Hire assessments are more commonly used in some hiring situations. It may also be helpful to collect more detailed data during the recruitment process on whether and why a search is canceled at various stages of the recruitment process.

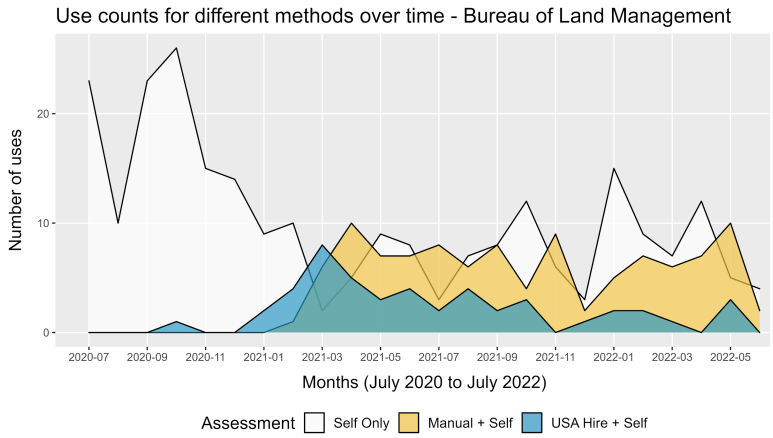
For applicant-level outcomes, further exploration of applicant characteristics could provide a better understanding of which candidates move through the key stages of the recruitment process. For instance, describing candidate characteristics among those who are designated as qualified but not referred to a cert list, those who choose to enter on duty among the selected, and those who remain in the position for some period of time in the future may provide a more complete picture of candidate characteristics through the recruitment process.

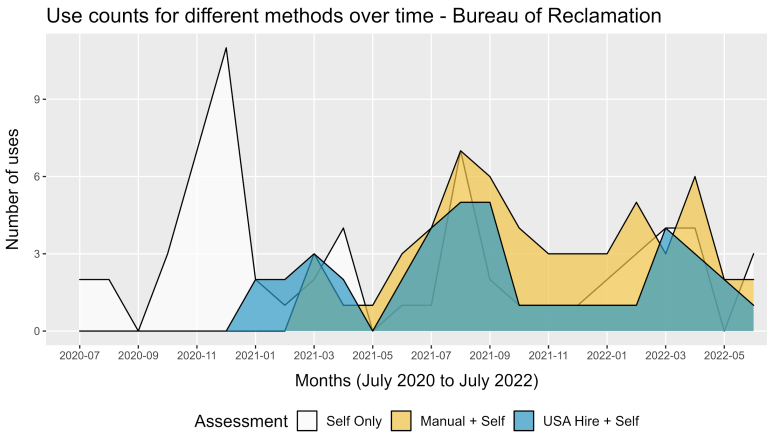
It may be possible to develop a randomized evaluation of the effects of different assessment methods on priority hiring outcomes. This type of study could provide insights into whether any differences in hiring outcomes are likely due to the use of one assessment method over another, rather than being driven by hiring officials' preferences or other parts of the hiring process. A first step in exploring a randomized evaluation may be to learn *why* hiring officials select different assessment methods in different circumstances (e.g., in different bureaus, for different job series) and to what extent hiring managers have the ability to use one assessment method over another.

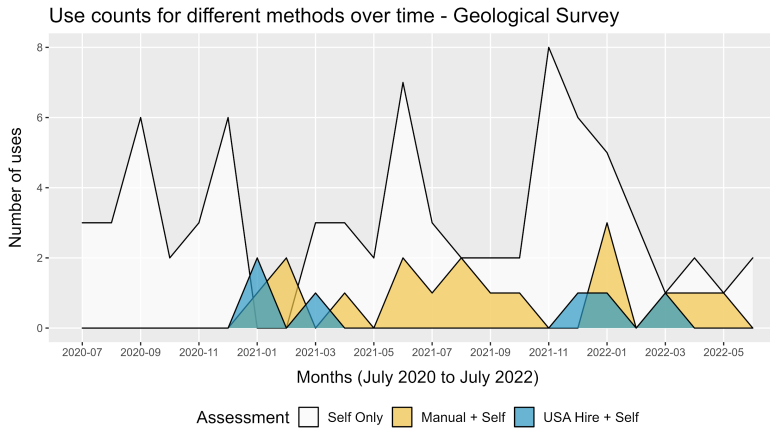
Appendix A

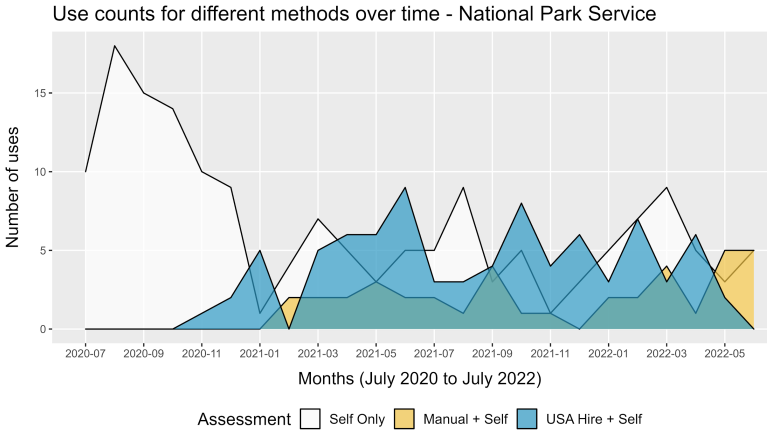
To describe trends in candidate assessments implementation across DOI, we illustrated utilization for each of the three assessment types (self-assessment only, self plus manual assessment, and self plus USA Hire assessment) over time using ridgeline plots. Three analyses examined 1) the use of the three assessment types across DOI, 2) the use of the three assessment types for each of the five DOI bureaus included in the study, and, 3) the use of the three assessment types for the 10 job series included in the study, each by month. Figure 2 and 9- illustrate the distribution of assessment utilization within DOI, across DOI's five largest bureaus and 10 job series from July 2020 to July 2022.

Figure A1. Implementation of Hiring Assessments for five DOI Bureaus, by month.









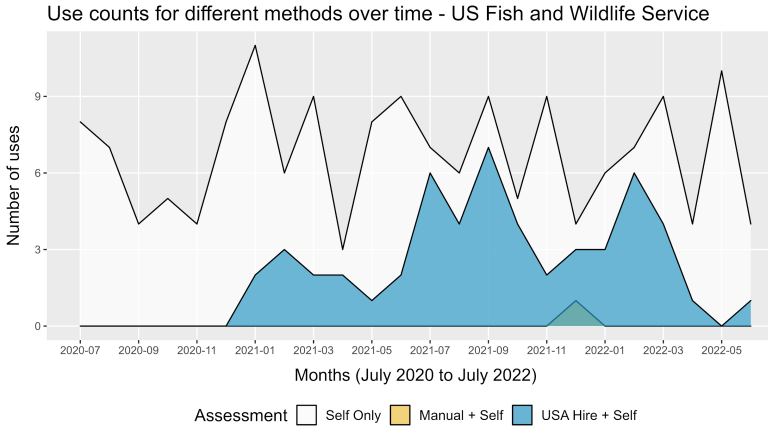
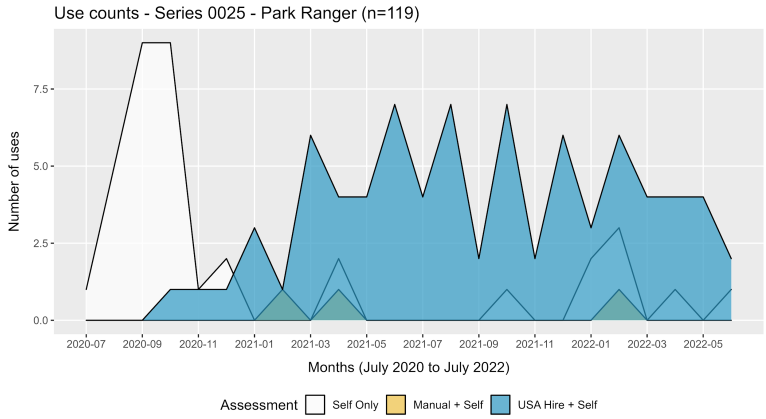
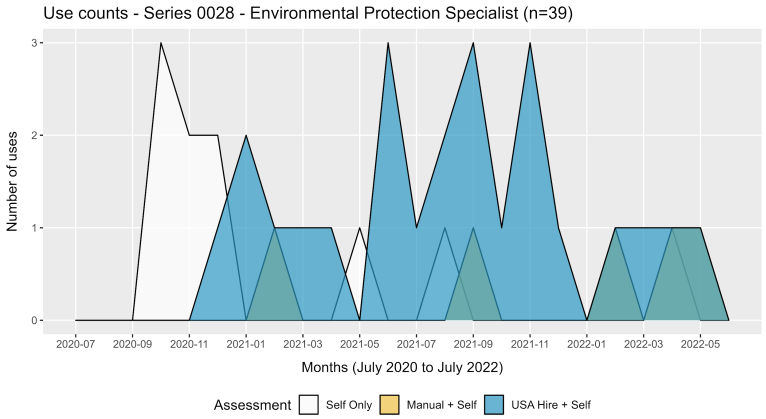
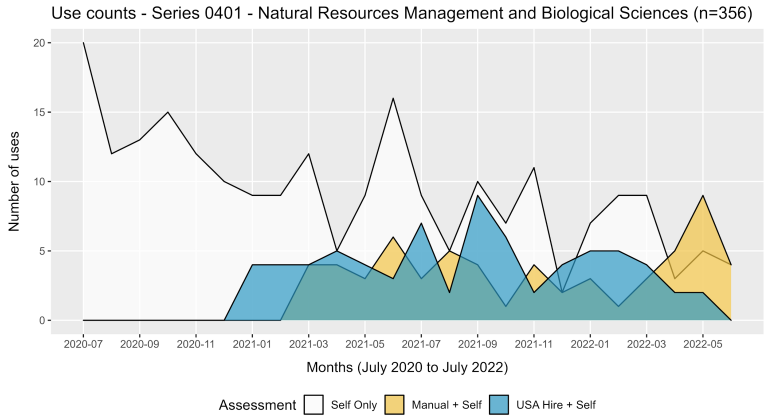
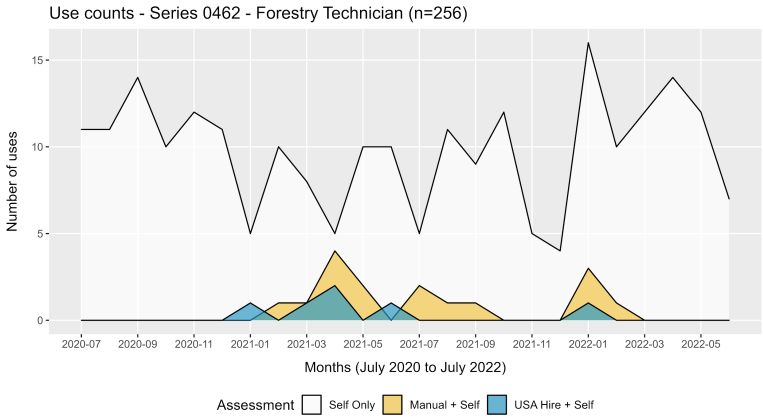


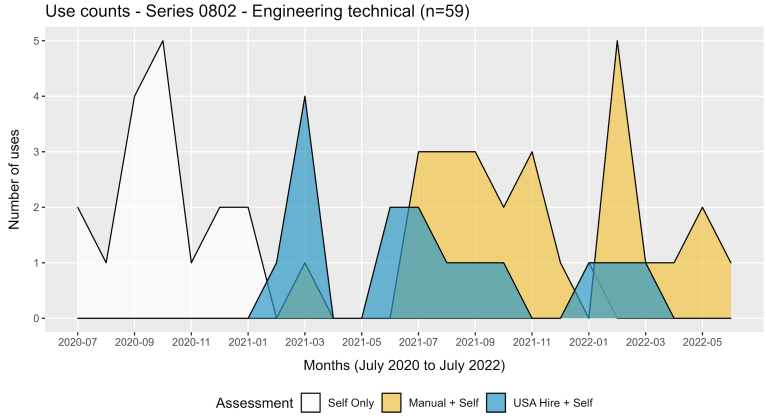
Figure A2. Implementation of Hiring Assessments for ten job series, by month.

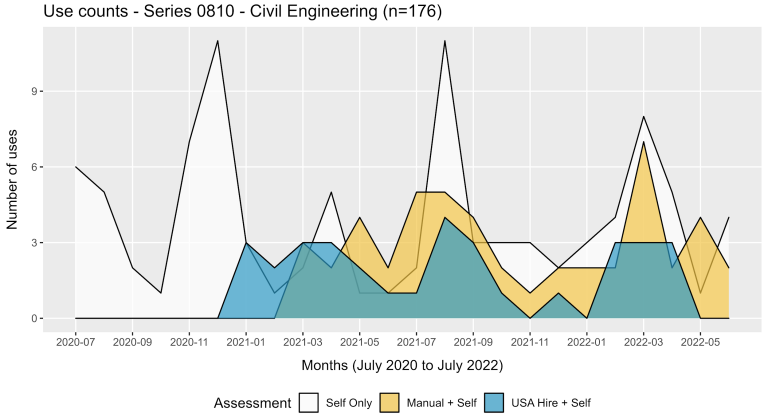


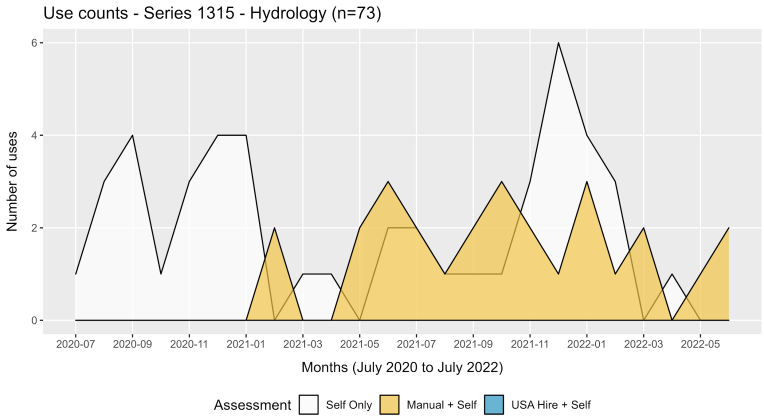


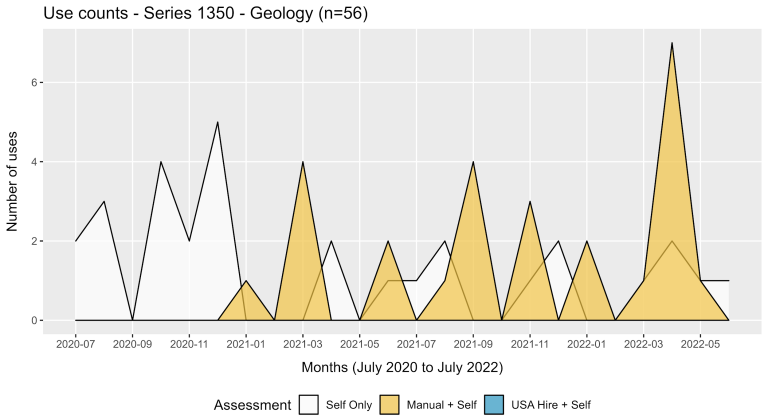


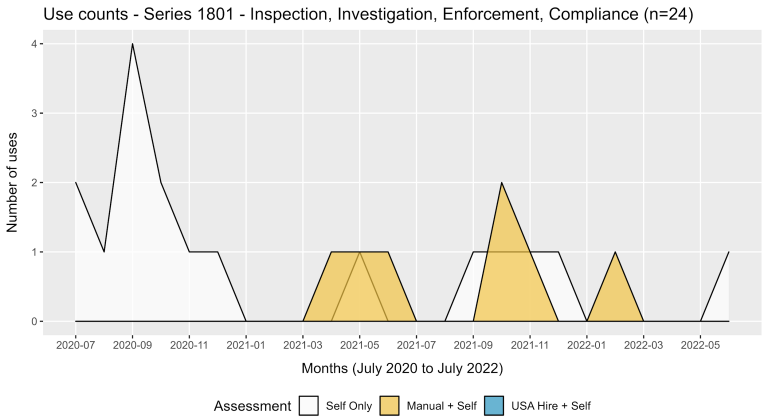


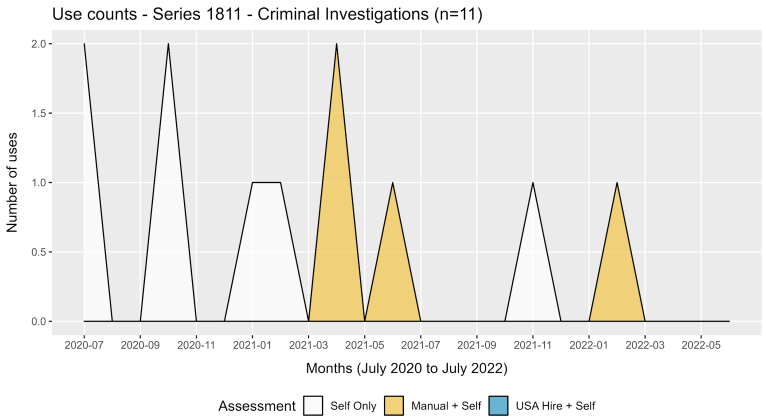





Use counts - Series 1315 - Hydrology (n=73)
Number of uses
0
2
4
6
2020-07
2020-09
2020-11
2021-01
2021-03
2021-05
2021-07
2021-09
2021-11
2022-01
2022-03
2022-05
Months (July 2020 to July 2022)
Assessment
Self Only
Manual + Self
USA Hire + Self


Use counts - Series 1350 - Geology (n=56)
Number of uses
0
2
4
6
2020-07
2020-09
2020-11
2021-01
2021-03
2021-05
2021-07
2021-09
2021-11
2022-01
2022-03
2022-05
Months (July 2020 to July 2022)
Assessment
Self Only
Manual + Self
USA Hire + Self


Use counts - Series 1801 - Inspection, Investigation, Enforcement, Compliance (n=24)
Number of uses
0
1
2
3
4
2020-07
2020-09
2020-11
2021-01
2021-03
2021-05
2021-07
2021-09
2021-11
2022-01
2022-03
2022-05
Months (July 2020 to July 2022)
Assessment
Self Only
Manual + Self
USA Hire + Self


Use counts - Series 1811 - Criminal Investigations (n=11)
Number of uses
0.0
0.5
1.0
1.5
2.0
2020-07
2020-09
2020-11
2021-01
2021-03
2021-05
2021-07
2021-09
2021-11
2022-01
2022-03
2022-05
Months (July 2020 to July 2022)
Assessment
Self Only
Manual + Self
USA Hire + Self

Appendix B

To explore how the hiring action-level outcomes differ by assessment method within bureaus, we conducted one-way ANOVAs for certification time, post-certification time, and recruitment success for each of the five bureaus as an exploratory analysis. If an F-test rejects the null hypothesis that the outcome did not differ across assessment methods, we conduct a Tukey's post-hoc analysis of mean differences and test whether the pairwise means are different.

Table B1. One-way ANOVA for Bureau of Land Management (BLM) hiring action-level outcomes

Outcome	***F statistic (df = 2)***	***p-value***	**Tukey's post-hoc analysis of pair-wise mean differences in outcomes**		
			Self vs. Manual	*USA Hire vs. Manual*	*USA Hire vs. Self- Report*
Certification time (days)	26.58	<0.001	-12.32 (p <0.001)	-7.33 (p = 0.012)	4.99 (p = 0.078)
Post-certification time (days)	2.00	0.138	*n/a*	*n/a*	*n/a*
Recruitment success (%)	2.03	0.132	*n/a*	*n/a*	*n/a*

Note: We reject the hypothesis of no differences in the outcome across assessment methods if the p-value from the F-test is less than 0.05.

Table B2. One-way ANOVA for Bureau of Reclamation (BOR) hiring action-level outcomes

Outcome	***F statistic (df = 2)***	***p-value***	**Tukey's post-hoc analysis of pair-wise mean differences in outcomes**		
			Self vs. Manual	*USA Hire vs. Manual*	*USA Hire vs. Self- Report*
Certification time (days)	13.57	<0.001	-10.10 (p <0.001)	-9.35 (p <0.001)	0.75 (p = 0.941)
Post-certification time (days)	1.11	0.338	*n/a*	*n/a*	*n/a*
Recruitment success (%)	5.07	0.007	0.022 (p = 0.966)	-0.271 (p = 0.022)	-0.294 (p = 0.009)

Note: We reject the hypothesis of no differences in the outcome across assessment methods if the p-value from the F-test is less than 0.05.

Table B3. One-way ANOVA for U.S. Geological Survey (USGS) hiring action-level outcomes

Outcome	***F statistic (df = 2)***	***p-value***	**Tukey's post-hoc analysis of pair-wise mean differences in outcomes**		
			Self vs. Manual	*USA Hire vs. Manual*	*USA Hire vs. Self-Report*
Certification time (days)	11.62	<0.001	-30.57 (p <0.001)	-40.44 (p = 0.003)	-9.87 (p = 0.618)
Post-certification time (days)	2.12	0.131	*n/a*	*n/a*	*n/a*
Recruitment success (%)	5.47	0.006	0.337 (p = 0.004)	0.304 (p =0.217)	-0.033 (p = 0.977)

Note: We reject the hypothesis of no differences in the outcome across assessment methods if the p-value from the F-test is less than 0.05.

Table B4. One-way ANOVA for National Park Service (NPS) hiring action-level outcomes

Outcome	***F statistic (df = 2)***	***p-value***	**Tukey's post-hoc analysis of pair-wise mean differences in outcomes**		
			Self vs. Manual	*USA Hire vs. Manual*	*USA Hire vs. Self-Report*
Certification time (days)	15.28	<0.001	-22.55 (p <0.001)	-12.84 (p =0.018)	9.71 (p =0.006)
Post-certification time (days)	0.40	0.670	*n/a*	*n/a*	*n/a*
Recruitment success (%)	3.38	0.036	0.175 (p = 0.070)	0.219 (p = 0.030)	0.044 (p = 0.742)

Note: We reject the hypothesis of no differences in the outcome across assessment methods if the p-value from the F-test is less than 0.05.

Table B5. One-way ANOVA for U.S. Fish and Wildlife Service (FWS) hiring action-level outcomes

Outcome	***F statistic (df = 2)***	***p-value***	**Tukey's post-hoc analysis of pair-wise mean differences in outcomes**		
			Self vs. Manual	*USA Hire vs. Manual*	*USA Hire vs. Self- Report*
Certification time (days)	9.40	<0.001	-20.60 (p =0.062)	-15.15 (p =0.222)	5.45 (p <0.001)
Post-certification time	0.39	0.679	*n/a*	*n/a*	*n/a*

(days)					
Recruitment success (%)	2.85	0.060	*n/a*	*n/a*	*n/a*

Note: We reject the hypothesis of no differences in the outcome across assessment methods if the p-value from the F-test is less than 0.05.

Appendix C

As an exploratory analysis we tabulated the representation of applicants with different characteristics – women, applicants of color (non-White or Hispanic), and Veterans with preference – for each of the five bureaus. These tabulations are calculated for each stage of the recruitment process by assessment method. We do not conduct statistical tests of the differences in representation by bureau.

Table C1. Proportion of female, non-White or Hispanic, and Veteran applicants to BLM vacancies, by assessment method and recruitment stage

a. Self-reported assessments			
	Applied *n* = 30,293	Made Cert-list *n* = 14,164	Selected *n* = 2,270
Female	0.23	0.199	0.195
Missing gender	0.338	0.325	0.342
Non-White or Hispanic	0.214	0.2	0.184
Missing race or ethnicity	0.343	0.331	0.352
Veteran	0.026	0.04	0.015
b. Manual assessments			
	Applied *n* = 2,540	Made Cert-list *n* = 703	Selected *n* = 118
Female	0.17	0.162	0.215
Missing gender	0.313	0.316	0.331
Non-White or Hispanic	0.28	0.242	0.154
Missing race or ethnicity	0.324	0.33	0.339
Veteran	0.143	0.192	0.161
c. **USAHire assessments**			
	Applied *n* = 3,727	Made Cert-list *n* = 690	Selected *n* = 78
Female	0.367	0.35	0.396
Missing gender	0.29	0.275	0.321

Non-White or Hispanic	0.171	0.14	0.115
Missing race or ethnicity	0.297	0.287	0.333
Veteran	0.055	0.135	0.077

Table C2. Proportion of female, non-White or Hispanic, and Veteran applicants to BOR vacancies, by assessment method and recruitment stage

a. Self-reported assessments			
	Applied *n* = 2,715	Made Cert-list *n* = 877	Selected *n* = 90
Female	0.312	0.255	0.3
Missing gender	0.344	0.342	0.444
Non-White or Hispanic	0.219	0.193	0.122
Missing race or ethnicity	0.352	0.356	0.456
Veteran	0.031	0.044	0.044
b. Manual assessments			
	Applied *n* = 1,098	Made Cert-list *n* = 326	Selected *n* = 60
Female	0.316	0.297	0.452
Missing gender	0.332	0.288	0.3
Non-White or Hispanic	0.228	0.222	0.19
Missing race or ethnicity	0.346	0.31	0.3
Veteran	0.085	0.132	0.1
c. USAHire assessments			
	Applied *n* = 941	Made Cert-list *n* = 197	Selected *n* = 23
Female	0.32	0.247	0.312
Missing gender	0.312	0.259	0.304
Non-White or Hispanic	0.285	0.297	0.188

Missing race or ethnicity	0.322	0.264	0.304
Veteran	0.058	0.183	0.043

Table C3. Proportion of female, non-White or Hispanic, and Veteran applicants to USGS vacancies, by assessment method and recruitment stage

a. Self-reported assessments			
	Applied *n* = 12,835	Made Cert-list *n* = 3,215	Selected *n* = 237
Female	0.481	0.483	0.575
Missing gender	0.312	0.252	0.266
Non-White or Hispanic	0.195	0.172	0.112
Missing race or ethnicity	0.323	0.264	0.287
Veteran	0.019	0.038	0.034
b. Manual assessments			
	Applied *n* = 1,354	Made Cert-list *n* = 255	Selected *n* = 21
Female	0.399	0.384	0.556
Missing gender	0.301	0.275	0.143
Non-White or Hispanic	0.191	0.137	0
Missing race or ethnicity	0.311	0.286	0.143
Veteran	0.026	0.071	0.048
c. USAHire assessments			
	Applied *n* = 1,259	Made Cert-list *n* = 377	Selected *n* = 20
Female	0.531	0.539	0.286
Missing gender	0.288	0.252	0.3
Non-White or Hispanic	0.171	0.106	0
Missing race or ethnicity	0.291	0.252	0.3

Veteran	0.006	0.019	0.05

Table C4: Proportion of female, non-White or Hispanic, and Veteran applicants to *NPS* vacancies, by assessment method and recruitment stage.

a. Self-reported assessments			
	Applied *n* = 10,842	Made Cert-list *n* = 2,471	Selected *n* = 214
Female	0.33	0.31	0.368
Missing gender	0.3	0.257	0.238
Non-White or Hispanic	0.224	0.195	0.151
Missing race or ethnicity	0.308	0.266	0.257
Veteran	0.045	0.115	0.117
b. Manual assessments			
	Applied *n* = 1,927	Made Cert-list *n* = 538	Selected *n* = 44
Female	0.376	0.392	0.576
Missing gender	0.33	0.279	0.25
Non-White or Hispanic	0.228	0.175	0.129
Missing race or ethnicity	0.344	0.29	0.295
Veteran	0.07	0.086	0.068
c. USAHire assessments			
	Applied *n* = 10,813	Made Cert-list *n* = 2,442	Selected *n* = 215
Female	0.311	0.361	0.335
Missing gender	0.26	0.216	0.251
Non-White or Hispanic	0.234	0.147	0.212
Missing race or ethnicity	0.27	0.23	0.274
Veteran	0.065	0.195	0.158

Table C5. Proportion of female, non-White or Hispanic, and Veteran applicants to FWS vacancies, by assessment method and recruitment stage

a. Self-reported assessments			
	Applied *n* = 26,158	Made Cert-list *n* = 11,610	Selected *n* = 589
Female	0.343	0.304	0.353
Missing gender	0.299	0.275	0.244
Non-White or Hispanic	0.239	0.215	0.203
Missing race or ethnicity	0.307	0.284	0.255
Veteran	0.1	0.165	0.068
b. Manual assessments			
	Applied *n* = 13	Made Cert-list *n* = 7	Selected *n* = 1
Female	0.143	0	NA
Missing gender	0.462	0.714	1
Non-White or Hispanic	0	0	NA
Missing race or ethnicity	0.462	0.714	1
Veteran	0.077	0.143	0
c. USAHire assessments			
	Applied *n* = 9,012	Made Cert-list *n* = 1,515	Selected *n* = 108
Female	0.458	0.493	0.7
Missing gender	0.275	0.239	0.259
Non-White or Hispanic	0.193	0.167	0.2
Missing race or ethnicity	0.284	0.257	0.306
Veteran	0.048	0.143	0.065