Screened Out

The Impact of Digitized Hiring Assessments on Disabled Workers

Michal Luria
Matthew U. Scherer
Dhanaraj Thakur

Ariana Aboulafia
Henry Claypool
Wilneida Negrón

November 2024

CENTER FOR DEMOCRACY & TECHNOLOGY

The **Center for Democracy & Technology** (CDT) is the leading nonpartisan, nonprofit organization fighting to advance civil rights and civil liberties in the digital age. We shape technology policy, governance, and design with a focus on equity and democratic values. Established in 1994, CDT has been a trusted advocate for digital rights since the earliest days of the internet. The organization is headquartered in Washington, D.C. and has a Europe Office in Brussels, Belgium.

coworker.org

Coworker.org is a laboratory for workers to experiment with power-building strategies and win meaningful changes in the 21st-century economy. At Coworker, we invest in the brilliance of workers by hosting and promoting workplace petition campaigns, prototyping fresh ideas for wielding influence at work and across industries, researching answers to questions about working conditions, and leveraging our vast network of workers in a wide variety of industries to reveal new insights, analysis, and data about what's happening in our economy. We support the leadership and vision of working people to imagine, design, and create our collective future.

AAPD
American Association of People with Disabilities

The **American Association of People with Disabilities** (AAPD) is a national nonprofit organization that works to increase the political and economic power of people with disabilities. As a national disability-led and cross-disability rights organization, AAPD advocates for full civil rights for over 71 million Americans with disabilities by promoting equal opportunity, economic power, independent living, and political participation.

Screened Out

The Impact of Digitized Hiring Assessments on Disabled Workers

Authors

Michal Luria, Matthew U. Scherer, Dhanaraj Thakur, Ariana Aboulafia, Henry Claypool, Wilneida Negrón

Contents

Contents

Michal Luria, Matt Scherer, Dhanaraj Thakur, Ariana Aboulafia, Henry Claypool, Wilneida Negrón

Executive Summary

Companies have incorporated hiring technologies, including AI-powered assessments and other automated employment decision systems (AEDSs), into various stages of the hiring process across a wide range of industries. While proponents argue that these technologies can aid in identifying suitable candidates and reducing bias, researchers and advocates have identified multiple ethical and legal risks that these technologies present, including discriminatory impacts on members of marginalized groups. This study examines some of the impacts of modern computer-based assessments ("digitized assessments") — the kinds of assessments commonly used by employers as part of their hiring processes — on disabled job applicants.

The findings and insights in this report aim to inform employers, policymakers, advocates, and researchers about some of the validity and ethical considerations surrounding the use of digitized assessments, with a specific focus on impacts on people with disabilities.

Methodology

We utilized a human-centered qualitative approach to investigate and document the experiences and concerns of a diverse group of participants with disabilities. Participants were asked to complete a series of digitized assessments, including a personality test, cognitive tests, and an AI-scored video interview, and were interviewed about their experiences. Our study included participants who identified as low vision, people with brain injuries, autistic people, D/deaf and/or hard of hearing people, those with intellectual or developmental disabilities, and those with mobility differences. We also included participants with diverse demographic backgrounds in terms of age, race, and gender identity.

The study focused on two distinct groups: (1) individuals who are currently working in, or intend to seek, hourly jobs, and (2) attorneys and law students who have sought or are likely to seek lawyer jobs. By studying these groups, we aimed to understand potential impacts of digitized assessments on workers with roles that require different levels of education and experience.
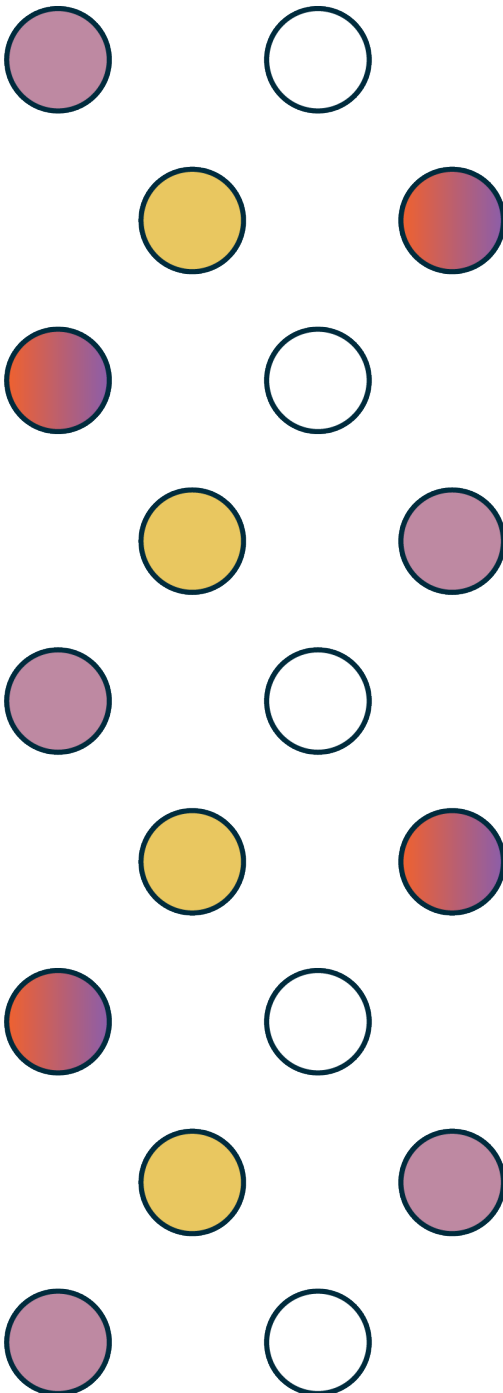
Findings

Disabled workers felt discriminated against and believed the assessments presented a variety of accessibility barriers. Contrary to the claims made by developers and vendors of hiring technologies that these kinds of assessments can reduce bias, participants commonly expressed that the design and use of assessments were discriminatory and perpetuated biases (*"They're consciously using these tests knowing that people with disabilities aren't going to do well on them, and are going to get self-screened out"*).

Participants felt that the barriers they grappled with stemmed from assumptions made by the designers in how assessments were presented, designed, or even accessed. Some viewed these design choices as potentially reflective of an intent to discriminate against disabled workers. One participant stated that it *"felt like it was a test of, 'how disabled are you?'"* Not only that, participants generally viewed the assessments as ineffective for measuring job-relevant skills and abilities.

Participants were split on whether these digitized assessments could be modified in a way that would make them more fair and effective. Some participants believed the ability to engage in parts of the hiring process remotely and asynchronously could be useful during particular stages, if combined with human supervision and additional safeguards. Most, however, did not believe that it would be possible to overcome the inherent biases against individuals with disabilities in how assessments are used and designed. As one participant put it *"We, as very flawed humans, are creating even more flawed tools and then trying to say that they are, in fact, reducing bias when they're only confirming our own already held biases."*

Given the findings of this study, employers and developers of digitized assessments need to re-evaluate the design and implementation of assessments in order to prevent the perpetuation of biases and discrimination against disabled workers. There is a clear need for an inclusive approach in the development of hiring technologies that accounts for the diverse needs of all potential candidates, including individuals with disabilities.

Michal Luria, Matt Scherer, Dhanaraj Thakur, Ariana Aboulafia, Henry Claypool, Wilneida Negrón

Recommendations

Below we highlight our main recommendations for developers and deployers of digitized assessments, based on participants' observations and experiences. Given the harm these technologies may introduce, some of which may be intractable, the following recommendations set out to reduce harm rather than eliminate it altogether.

Necessity of Assessments: Employers should first evaluate whether a digitized assessment is necessary, and whether there are alternative methods for measuring the desired skills with a lower risk of discrimination. If employers select to use digitized assessments, they should ensure that the assessments used are fair and effective; that they measure skills or abilities directly relevant to the specific job, and that they can do so accurately.

Accessibility: Employers must ensure assessments adhere to existing accessibility guidelines, like the Web Content Accessibility Guidelines (WCAG)[1] or initiatives of the Partnership on Employment and Accessible Technologies (PEAT)[2], and that the selected assessments accommodate and correctly assess the skills of disabled workers with various disabilities.

Implementation: For effective, fair, and accessible assessments, employers can take additional steps to potentially reduce biases by implementing significant human oversight in all assessment processes, using assessments to supplement, not replace, comprehensive candidate evaluations, and being transparent about when and how assessments are used.

---

1   https://www.w3.org/TR/WCAG21/
2   https://www.peatworks.org/

Introduction

"It was soul crushing [...] AI is great and all, but these are people's lives." – Reaction of an interviewee after completing a series of hiring assessments

The recent proliferation of artificial intelligence (AI) and other automated technologies and tools has permeated many companies' business practices and workflows, including recruitment and hiring processes. Successfully hiring employees is a challenging and time-consuming task, and many employers have turned to technology to automate parts of the process. Proponents argue that such automation can help with collecting, screening, and recommending job candidates. However, some candidates may be marginalized by this automation, including people with disabilities. In this report we focus on the impacts of computer-based assessments – specifically AI-scored video interviews, gamified personality evaluations, and cognitive tests – on disabled people.

The Integration of Technology in Hiring

Technology has been incorporated into nearly every stage of the hiring process (Rieke & Bogen, 2018), from targeting advertisements for jobs, to collecting and screening applications and conducting interviews. While AI-powered assessments and other automated employment decision systems (AEDSs) have drawn attention from both the media and advocates, modern hiring technologies can take many forms and can be used in many different ways. They include gamified tests, assessments that rely on facial recognition and analysis, and computerized or algorithmic versions of assessments that have long been used by employers in hiring processes (like personality tests, cognitive tests, and more) (Mimbela & Akselrod, 2024).

While the lack of transparency regarding companies' use of technology in hiring and the lack of regulation of such technologies make it hard to precisely quantify the prevalence of these hiring technologies, various surveys and studies indicate that their use is widespread. For example, the chair of the Equal Employment Opportunity Commission suggested that "some 83% of employers, including 99% of Fortune 500 companies, now use some form of automated tool as part of their hiring process" (Hsu, 2023). Another survey noted that 76% of companies
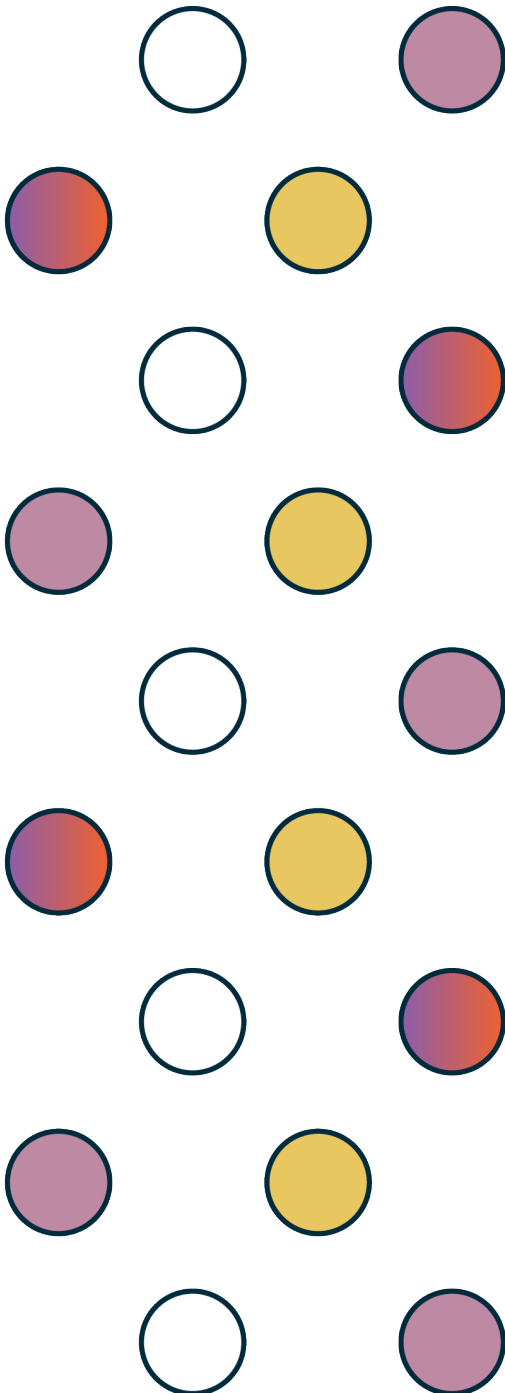
with more than 100 employees use personality tests, and that employers are turning to algorithms to administer and analyze the tests at a larger scale (Brown et al., 2020).

The developers and vendors of AI-integrated hiring technologies claim that their tools can help employers identify the applicants that are the best fit for a given job, help sort and organize candidates (ACLU, 2024), and potentially even reduce bias in the hiring process (Savage & Bales, 2016; Raghavan et al., 2020). In contrast to the claims of vendors, research shows that the use of modern hiring technologies can introduce a range of ethical and legal risks (Rieke & Bogen, 2018), including privacy risks (Kim & Bodie, 2020) and a high risk of enabling employment discrimination based on race (Gershgorn, 2018), gender (Dastin, 2018), disability (Brown et al., 2020; Glazko et al., 2024), and other characteristics, including through perpetuating implicit bias (Persson, 2016).

In addition to discrimination concerns, the use of these technologies also raises questions around effectiveness. Many modern computer-based employment assessments assess skills or traits that are not necessary for some jobs (Akselrod & Venzke, 2023). For example, personality assessments measured general traits like positivity, emotional awareness, and liveliness (ACLU, 2024). Such characteristics are not clearly linked to most job functions, and risk screening out workers with autism or mental health conditions, like depression and anxiety.

Further, these kinds of assessments may not be able to meaningfully measure or predict the skills and qualities they purport to assess in the first place (Stark et al., 2021; Birhane, 2022). For example, recent research shows the validity of cognitive ability tests for predicting future job performance ratings has been substantially overestimated for several decades (Sackett et al., 2022). It has also long been established that cognitive ability tests often have adverse impacts based on race (Outtz & Newman, 2009; Cottrell et al., 2015).

The Impact of Hiring Technology on Disabled Workers

Job seeking has long been a process riddled with barriers for people with disabilities for a number of reasons, including ableist norms about desired qualities of a worker, choices disabled people have to make about whether to disclose their disabilities, and approaches to evaluating and communicating with applicants that don't account for their disabilities (Fruchterman & Mellea, 2018; Bonaccio et al., 2020). Despite significant gains since the passage of the Americans with Disabilities Act in 1991, the

participation rate of disabled individuals within the labor force is approximately half that of non-disabled individuals, and the unemployment rate for disabled workers is roughly double that of non-disabled workers (National Trends in Disability Employment, 2024; Bureau of Labor Statistics, 2024).

The use of modern hiring technologies, including those that use AI as part of the assessment or scoring process, may create even more barriers. AI systems often fail to account for the needs, experiences, and perspectives of disabled people (Brown et al., 2020; Williams, 2024). For example, a recent study found that résumé sorters incorporating OpenAI's GPT-4 exhibited prejudice in rankings if they contained activities or awards suggesting the candidate had a disability (Glazko et al., 2024). These and other issues can lead to a variety of negative consequences for disabled workers (Fruchterman & Mellea, 2018; Bonaccio et al., 2020).

Some of these barriers stem from the fact that disabled workers' needs are often overlooked in the design and evaluation of selection procedures, both those that leverage automation or AI and those that do not (Papinchock et al., 2023). This is especially likely to happen when hiring procedures and technologies are designed without input from disabled workers and disability experts, and thus do not consider the full range of people who may use a new technology or feature (Brown et al., 2020). Examples include systems that rely on facial recognition or automated analysis of interactions with a computer (Rieke & Bogen, 2018), or automated systems designed to recognize and analyze speech. These kinds of systems are commonly used to power video interviewing systems in hiring, and have been shown to perform worse for speakers with a variety of disabilities (Tu et al., 2016; Glasser et al., 2017; Hidalgo Lopez et al., 2023).

While there is clear evidence of the harms certain hiring technologies can have, including for disabled workers, there is a need for more research examining the extent and nature of hiring technologies' impacts on jobseekers with disabilities. In particular, we identified a gap in research regarding the multi-faceted experiences of disabled workers in engaging with modern hiring technology, as well as in understanding how hiring tools may have different impacts on job seekers with different kinds of disabilities. In this research report we contribute to addressing this gap by examining the experiences of people with disabilities with certain kinds of digitized assessments. In particular, the focus of this report is on the impact of computer-based hiring assessments — including personality testing, cognitive tests, and an AI-scored video interview (hereafter referred to collectively as "digitized assessments") — on disabled workers.[3]

---

3   Under definitions used by the United States Department of Labor and the Census Bureau, a worker (i.e., a member of the labor force) is someone who is either working or who is actively seeking employment.

Michal Luria, Matt Scherer, Dhanaraj Thakur, Ariana Aboulafia, Henry Claypool, Wilneida Negrón

Developers of these tools and employers who use them generally push back against the need for regulation of these technologies, often citing their own internal studies to claim that their hiring technologies do not introduce or exacerbate bias. Further, disability discrimination is an underemphasized area in AI governance and in shaping AI policy. Our goal is thus to provide findings and recommendations that can not only inform the research community and broader public, but that can also be useful for policy-making, including providing policymakers with additional evidence of the harms modern hiring technologies can have on diverse individuals with disabilities.

Research Approach

Engagement with impacted communities is a key element in detecting and analyzing potential discriminatory impacts of hiring technologies. To that end, this study adopts a human-centered, qualitative research approach, in which we asked a group of diverse participants with disabilities to attempt a range of digitized assessments, and subsequently interviewed them. In those interviews, we asked participants to reflect on their experiences and concerns through a disability-centered lens.

Participants

The study included 17 participants who completed (or attempted to complete) all studied assessments and completed a final interview. Participants had a diverse range of disabilities, and many had multiple types of disabilities (see Appendix 1 for full list). The most commonly reported disabilities among participants were cognitive disabilities (seven participants), vision disabilities (seven participants), and ambulatory disabilities (six participants).

We also aimed to include a demographically diverse group of individuals in the study: Participants' ages ranged between 18-59, and a total of eight self-identified as white, three as Black, three as Hispanic or Latino, one as Pacific Islander, one as Middle Eastern, and one as multi-racial. Eight participants self-identified as women, five as men, two as non-binary, one as genderqueer, and one identified as a woman and as non-binary.

Participants recruited in the study belonged to one of two groups: individuals who have sought or are likely to seek hourly jobs, and attorneys or law students who have sought or are likely to seek lawyer jobs. We refer to these groups as "hourly workers" and "lawyers" throughout the report. Of the 17 participants, 11 were hourly workers and six were lawyers.

We studied these two groups in order to gain perspectives on the potential impacts of digitized assessments on workers with roles that require no prior educational credentials, as well as the potential impacts on workers with roles that required a college degree and specialized education and/or experience. We selected lawyers for the higher-education group because law firms are already using some of the hiring
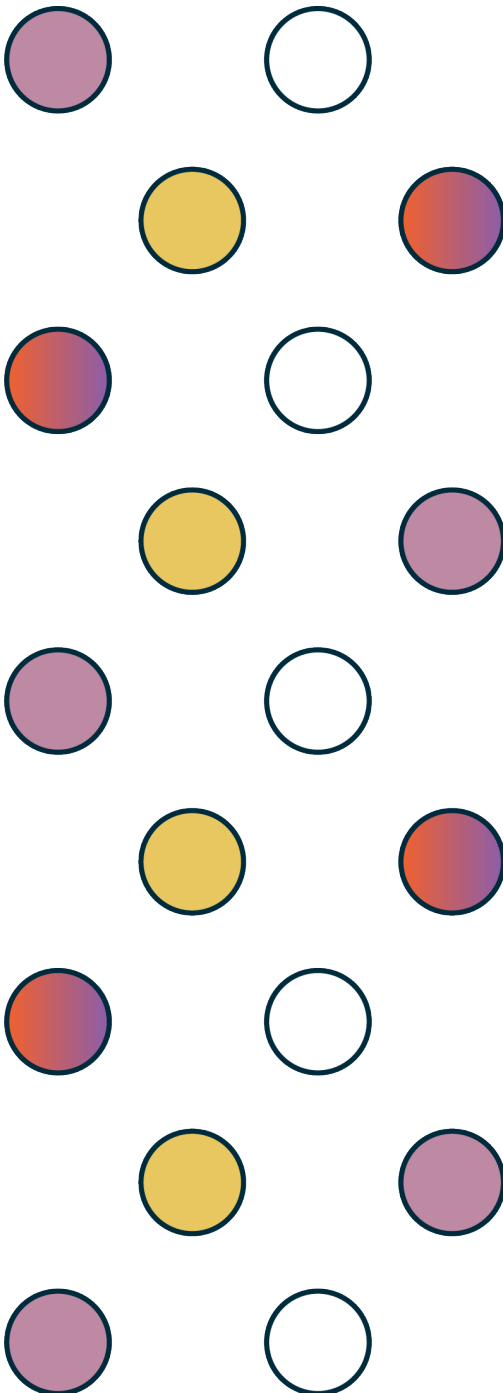
tools included in the study, based on a review of the marketing materials of several hiring technology vendors as well as a review of the test preparation websites that were used in the study. The two groups provide a range of expertise and experiences – given the qualitative approach and the small sample, we did not seek to identify systematic differences between the two groups.

Procedure

Selected participants were scheduled to complete three parts of the study: Two assessment sessions followed by one semi-structured interview. The assessment sessions were completed on two test preparation platforms intended to simulate digitized assessments that job seekers may encounter in a hiring process. Participants were asked to complete a set of up to eight simulated hiring assessments, guided by a CDT researcher.[4] Because participants completed the assessments on test preparation platforms, they received scores after each assessment (in contrast, when actually applying for jobs, applicants generally do not receive information about how they performed on assessments they are asked to take). That said, some participants noted that they either did not see or were not interested in viewing their scores.

The set of assessments for the study consisted of a combination of personality and work behavior tests that attempted to assess a range of specific skills (attention, risk taking behavior, etc.), along with gamified cognitive and personality tests, an emotional intelligence test, and a video interview that included automatic analysis of participants' facial expressions. For more information on the selected assessments, see Appendix 2.

Participants were then scheduled for a remote, one-hour semi-structured interview in which they were asked to reflect on their experiences and perceptions of the tests and their potential use in a hiring process. All assessments and interviews were completed between November 2023 and May 2024.

The results in this report are primarily based on a qualitative thematic analysis of interviews with participants, in which they were asked about their experiences going through the digitized assessments and their perspectives on the assessments, as well as a basic statistical analysis of participants' scores on the assessments. Thematic analysis allowed us to identify study participants' experiences, meanings, and realities, and to report on the patterns that people with a range of disabilities reported.

---

4   CDT's researcher assisted participants in finding each test on the platform, but not with the instructions or completion, as our goal was to examine, among other things, its accessibility and ease of use.

Limitations

The findings have some limitations, due to the study focus and design. First, they intend to qualitatively examine the impact of assessments on a group with a range of disabilities, and not intended to be generalizable for the entire group. Examining assessment scores using a larger sample that allows for both more robust statistical analysis and disaggregated analysis for participants with different types of disabilities is an important area of future work.

Second, in the study we used assessment platforms by test preparation companies. While the test preparation companies have related entities that market tests to employers for real-world candidate assessment too, we could not validate whether those employer-facing companies' assessments were exact duplicates of their test preparation counterparts.

Lastly, we do not have knowledge about how exactly employers use assessment scores (although prior work suggests they are used for screening (Hunkenschroer & Luetge, 2022)), or the duration and intensity of tests that they ask job seekers to complete. Different test sequences, or sequences that are shorter than our 2-hour sessions may have other results.

Findings

This section describes the findings from participants' reflection on their engagement with the selected digitized assessments. These insights may not generalize to all digitized assessments, but reveal critical issues that may apply to many. Further, in the interviews, participants frequently struggled to distinguish between different assessments, and rather provided insight into the collective experience of engaging with digitized assessments in a recruitment process.

At several points in the discussion below, we describe applicants' scores on the assessments, connecting their reflections on the experience of taking the assessments with their quantitative performance on the tests. In some areas of this discussion, we highlight the differential impacts of particular tests on participants with specific types of disabilities.

1. Digitized Assessments Disadvantage Disabled Workers

1.1 Participants felt that most of the digitized assessments were discriminatory and perpetuated biases

One of the most common observations among participants after completing the assessments was that they felt the tests were discriminatory. Participants perceived the potential for discrimination in almost every tested assessment, but frequently struggled to point out which ones were more or less discriminatory. Rather, as noted above, participants tended to perceive the assessments as a collective, overwhelming barrier to their success.

This barrier may be due to design and presentation assumptions made by the assessment platforms, and a lack of consideration of diverse user groups. However, several participants viewed this kind of discrimination not as an unintended consequence of employers using digitized assessments, but as a way to intentionally exclude disabled workers from the hiring process:
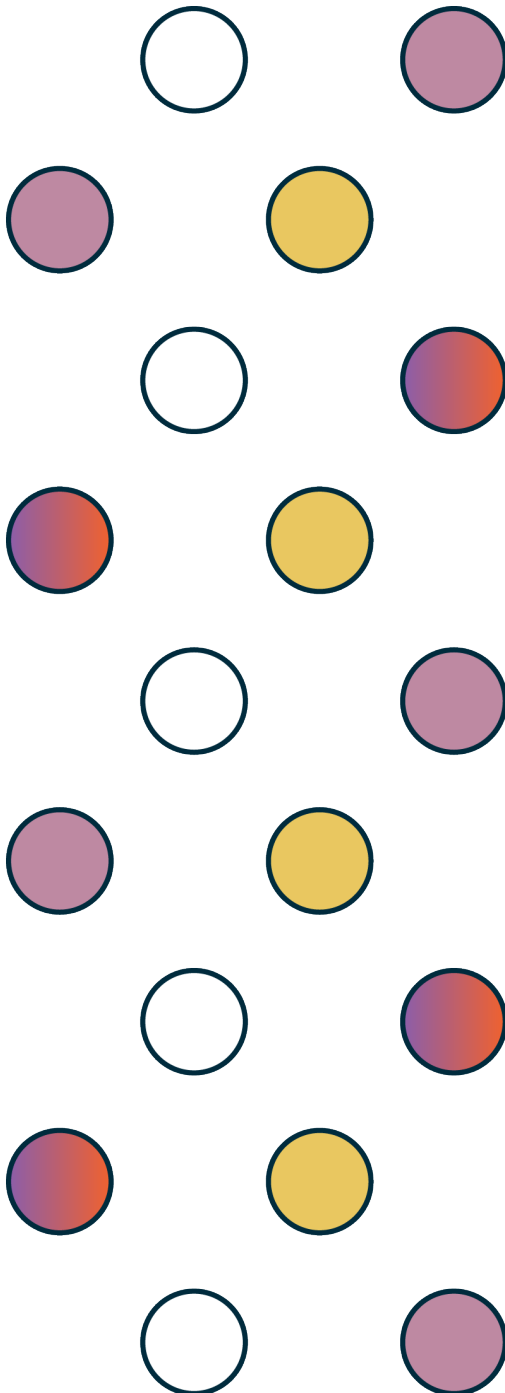
"They're consciously using these tests knowing that people with disabilities aren't going to do well on them and are going to get self-screened out, either because they don't finish the tests or because they do so horribly" (participant with a brain injury).

"These tests can be used as a pretext. 'I don't want to hire this candidate. Let me give him the test and see how he does.' 'Oh, he failed. He didn't get that much of a score.' 'Okay. Well, I don't want to hire him because of these scores.' There's a lawsuit. 'You didn't hire me because I'm disabled.' 'No, here's the test. You failed it'" (participant with a vision disability).

The convergence in participants' perceptions of the outcomes as inherently biased and discriminatory stands in contrast to the purported potential of these tools to reduce bias, a view that many developers and vendors of hiring technologies promote.

The convergence in participants' perceptions of the outcomes as inherently biased and discriminatory stands in contrast to the purported potential of these tools to reduce bias, a view that many developers and vendors of hiring technologies promote (Savage & Bales, 2016). When we asked participants about the assessments' potential to reduce bias against persons with disabilities, most did not see how these digitized assessments would do so. Not only that, some participants characterized the assessments' design and their underlying assumptions as ableist, that only compound existing discrimination and biases:

"We, as very flawed humans, are creating even more flawed tools and then trying to say that they are, in fact, reducing bias when they're only confirming our own already held biases, and then that just creates a snowball effect of, 'Oh, well, I guess disabled people just aren't good for the workforce,' because our completely unbiased tools are telling us that they aren't" (participant with cognitive and ambulatory disabilities).

A few participants did suggest that it was possible for the use of the assessments to reduce bias, but only in situations when their disability does not impact their performance, and thus does not reveal one's disability. However, as is later discussed, it is likely that the use of assessments will reveal disabilities, whether participants are ready for that or not.

1.2 Digitized assessments presented a variety of accessibility barriers to people with disabilities

Another major concern raised by participants was the lack of accessibility features, which prevented them from completing the tests to the best of their abilities, or even at all. Many aspects of the tests contributed to the lack of accessibility, including their visual characteristics, the layout of information, and the format of assessments.

Michal Luria, Matt Scherer, Dhanaraj Thakur, Ariana Aboulafia, Henry Claypool, Wilneida Negrón

Most participants did not see any accessibility accommodation options, or assumed there weren't any. Those that explicitly looked for information about accommodations reported that they could not find any information about accommodations, including information about requesting them. Indeed, one platform had only one setting for requesting additional 25% more time on the test preparation platform, while the other did not offer any specific accommodations.

In attempting to complete tasked assessments, most participants completed them as best they could, but in some cases participants had to skip a test. Below, we expand on participants' experiences of accessibility barriers, broken down by design choice and its impact.

Choice of Graphics

One of the main accessibility challenges originated in the use of color, whether for background use or on specific graphical elements (such as buttons). This use of color often created challenges for low vision participants, or those who were colorblind, and as noted earlier, participants reported no accessibility features that addressed this issue:

> *"The games that you had to play, the description [text on screen] and the background color made it very difficult to see. I'm colorblind [...] it was really difficult and there was no option to change the color to something that I could access [...] That was an additional barrier" (participant with a vision disability).*

The problem was not just in a single test or in the gamified tests in general (although the gamified interfaces tended to be particularly poor in terms of accessibility) — all of the assessments included graphical elements that were inaccessible to varying degrees. In fact, most participants who were blind or low-vision reported that their screen readers could not interpret the assessments. As a result, participants with screen readers were forced to guess, for example, if they were clicking in the right spot, or to spend time attempting to confirm that they'd selected their intended answers:

> *"The lack of auditory feedback for a screen reader user meant that things were changing dynamically in a way that made me concerned that I was actually either missing entire prompts or answering them incorrectly. Then that resulted in a lot of backtracking and duplicative effort" (participant with a vision disability).*

The problems posed by assessments with inaccessible graphical elements are also evident in participants' scores on a gamified arithmetic assessment that required participants to evaluate and click on dynamic graphic elements showing math equations on different areas of the screen, and on an attention assessment that required participants to evaluate whether dynamic illustrations were pointing left or right in rapid succession. Only one

The problem was not just in a single test or in the gamified tests in general... all of the assessments included graphical elements that were inaccessible to varying degrees.

of the participants with vision disabilities was able to complete the latter assessment, and only two were able to complete the gamified arithmetic assessment, getting some of the lowest percentile scores possible.[5]

Information Density

Another significant accessibility barrier, noted by multiple participants with cognitive disabilities, was that the assessments often required participants to process large amounts of information presented on a screen. Some of the information was also graphic-heavy, which added to people's cognitive load:

> *"The way that information was presented [across all assessments] was really dense. There was too much visual stimulation on a page. For some of it, I just had a hard time even with my visual processing problem, sorting out the junk from the actual meaningful content [...] if you have a visual processing issue, your brain still has to go through all of the different colors, the font changes, the lines, the separate sections"* (participant with a cognitive disability).

In some cases, the information was presented on multiple parts of the screen at the same time, requiring the participant to change their focus from one element to another. The additional information and graphical bloat were viewed not only as redundant but also overwhelming given that participants were trying to complete a test. Participants questioned whether all this information was even necessary:

> *"There's a bunch of minuscule details and numbers that are added to some of the questions that have nothing to do with how you're solving it. It's just extra information and makes it more confusing and more stuff to read on top of the fact you're timed"* (participant with cognitive and vision disabilities).

These issues were compounded by the fact that some of the tests were perceived as too long. One participant with a brain injury disability commented on the length of the "Situational Judgment" test, a test that asked participants how they would act in a range of professional settings, and consisted of 18 scenarios and four courses of action for each: *"I seriously wanted to scream and pull my hair out"* (participant with a cognitive disability). Five out of seven participants with cognitive disabilities were either not able to complete the situational judgment test or received scores in the bottom quartile (25%).[6]

---

5   Of the two participants with vision disabilities who completed the arithmetic gamified test, one of the participants scored in the 1st percentile and the other scored in the 3rd percentile.

6   Of the seven participants with cognitive disabilities, one was not able to complete the situational judgment test, three received percentile scores below ten, and the remaining three received percentile scores of 22, 67, and 72 respectively, where the range of possible percentile scores was 0-99.

Michal Luria, Matt Scherer, Dhanaraj Thakur, Ariana Aboulafia, Henry Claypool, Wilneida Negrón

Many described gamified assessments as *"ridiculous"* and *"silly."* But more importantly, they pointed out that the use of gamified assessments was visually distracting, added unnecessary information that added cognitive load and overall tended to be overwhelming.

Choice of Assessment Format

In the study, we selected several assessment types to examine how participants with disabilities experience diverse types of assessments for hiring. We found that some assessment formats were less accessible than others. One particularly concerning format was that of gamified interfaces – assessments that used game-like interfaces (e.g., asking users to inflate a balloon that eventually bursts) to assess candidates' skills and traits. Participants with a range of disabilities asked again and again: why games? Many described gamified assessments as "*ridiculous*" and "*silly*." But more importantly, they pointed out that the use of gamified assessments was visually distracting, added unnecessary information that added cognitive load and overall tended to be overwhelming. This critique pointed out some of the assumptions of using gamified assessments, as they require significantly more effort to process and may disadvantage people with disabilities, even though the same qualities could have been assessed in other ways (for example, by using text-based scenarios).

Another assessment format that participants described as particularly inaccessible was the "emotional intelligence" test, which asked participants to identify what emotions, such as anger, contempt, disgust, surprise, fear, or happiness, were being portrayed by people in a set of still images on the screen. This assessment, too, was inaccessible to people with vision disabilities, as well as for other participants. Two out of the 17 participants were not able to complete this test, and the average accuracy score for the 15 who were able to complete the test was 31.2 out of 100. But these low scores were not necessarily due to a lack of emotional intelligence skills – as prior work shows, assessments themselves tend to be inaccurate in evaluating autistic and other neurodivergent workers (Brown et al., 2020).

Choice of Language

Participants felt that the assessments assumed a particular cultural context and form of communication. One participant with a hearing disability shared:

> *"As a deaf person, I would prefer maybe an ASL [translation] because that's my primary language. Often, we have some surveys or tests like state tests at school, and they have to have a video of sign language doing the instructions because English, it's our second language. [...] If I would've had that option, I would've maybe understood [the instructions] more clearly."*

As this example highlights, assessments may be presented to applicants using a form of communication that is not only less-preferred or less accessible, but that in some cases may prevent entire groups of individuals from being able to successfully complete the assessment at all.

The choice of communication language would impact not only the participant's understanding of the instructions but also their process of providing responses — when discussing the interview assessment, this participant noted that they could more accurately express their tone when signing. However, the systems underlying the interview module may not be able to process signed responses accurately, if at all, and an applicant would likely not know whether a signed response could even be processed when deciding how to provide their answers.

Given the gaps in accessibility, some participants opted to improvise workaround techniques to make the assessments somewhat more accessible. One blind participant described how they used functionality designed for other purposes within their screen reader to get around the inaccessibility of a gamified test that required participants to inflate a virtual balloon by clicking on it:

> *"What I was doing to actually make it accessible, I was performing optical character recognition on the screen, and then simulating mouse clicks on the inflate and collect buttons as I went [...] It is a strategy that is – I'd say, an intermediate to advanced screen reader user technique. [...] Resorting to it is, I think, an indication that a page is not well- optimized" (participant with a vision disability).*

"...[It] should say, for instance, if you're visually disabled, we offer an audio version of the instructions. If you're dyslexic, we offer an audio version. If you are slow, if you have anxiety, and you require more time, well, click here."

As the participant pointed out, not everyone will have the skill to be able to set up that kind of workaround, and thus employers cannot assume such practices will improve accessibility of the assessments. Further, even if workarounds are possible, putting the onus on the applicant to find a way to make the assessment accessible creates a significant additional burden in time and effort that applicants without disabilities would not face. Discussion of web overlays that similarly manipulate the presentation of a website without altering its structural inaccessibility underscores that such participant-driven workarounds do not make assessments accessible (Harper, 2024).

Overall, participants highlighted the need for customization of interfaces across the board, given the range of experiences disabled people can have:

> *"Maybe a dropdown menu with large font and bright colors for the lettering and things like that. [It] should say, for instance, if you're visually disabled, we offer an audio version of the instructions. If you're dyslexic, we offer an audio version. If you are slow, if you have anxiety, and you require more time, well, click here." (participant with a hearing, vision and cognitive disability).*

Michal Luria, Matt Scherer, Dhanaraj Thakur, Ariana Aboulafia, Henry Claypool, Wilneida Negrón

2. Participants Expressed Concern About the Efficacy and Benefits of Digitized Assessments

2.1 The tested assessments were perceived as confusing, simplistic and irrelevant

Participants frequently pondered what the assessments were trying to measure, and how those presumably tested skills were relevant for the job. Some tests seemed more straightforward in their goal, such as the gamified mental arithmetic test. But other tests evoked confusion, skepticism and frustration given the resulting experience for people with some disabilities:

> *"I think [all assessments] were really measuring your visual processing speed and ability, even though that's not what I'm sure the test designers would say, but for someone with a disability trying to get through the test, it really felt like it was a test of, 'how disabled are you?'" (participant with a brain injury related disability).*

This point raises a larger concern about the use of standardized test formats and their impact on disabled people. Participants argued that these assessments reward those who are good at taking tests and are more prone to being biased against people with disabilities, especially as some will have different ways of processing information:

> *"I do have a general bias against standardized testing because they're not really neurodivergent or disability-friendly. When you look at test scores, because with tests and with quizzes, there are always going to be people who are just really good test takers" (participant with an ambulatory disability).*

> *"It's not going to be the best person for the job or the most capable person for the job, it's going to be who can beat the test [...] it's the way the test is designed. It penalizes you for processing information differently." (participant with a brain injury-related disability).*

A related outcome is that these assessments were perceived to overlook neurodiversity and reduce applicants to simple categories. Predefined answers from which test takers had to select (which was the case in almost all assessments) could not represent participants' thought processes or unique perspectives, and often suggested that there was a single right answer to each question. Ultimately, this can negatively impact both applicants and employers who may end up "*missing out on great candidates*" (participant with cognitive and ambulatory disabilities). Another participant summarized this broader impact on recruitment:

"...for someone with a disability trying to get through the test, it really felt like it was a test of, 'how disabled are you?'"

"If you do have a negative experience, it can completely turn you off from pursuing positions that require completion of these [tests]. That seriously impacts your ability to meet any diversity or greater reach that you might have been aiming for. That's something that could literally become a blind spot in your recruitment" (participant with a vision disability).

2.2 Digitized assessments were ineffective in assessing the intended skills

Most participants felt that the assessment questions, as well as their scores, were a poor representation of their skills and abilities, in addition to their outright discriminatory nature. For instance, autistic participants stated that the emotional intelligence test does not correctly reflect their ability to judge how people are feeling in social context. One participant explained:

"When you're just looking at a picture to judge someone's facial expression, it's just a guessing game, especially if you're somewhere on the autism spectrum [...] a lot of people with those disabilities will use the social context [in real situations] to guess how somebody's feeling or to have more information on how they're feeling. If somebody is unable to read facial expressions well, especially without context, this isn't a proper test of that at all" (autistic participant).

Another example is that most of the assessments had a set of predefined answers to select from. In some cases, like the Situational Judgment test, in which participants had to explain how they would act in a range of professional scenarios, participants expressed a desire to be able to provide more nuanced answers that explained their reasoning. The lack of this functionality was especially important for neurodiverse participants, as the predefined answers frequently did not represent their thought processes and perspectives, and often suggested that there was a single right answer to each question.

Some participants offered the experience and success of their respective careers as evidence for the inefficacy of the assessments given the low scores they received:

"I think [the assessments] were ineffective in accurately gauging my capabilities of recognizing and calculating things. It's troublesome because if one were to look at my résumé and my background and actually speak with me, they would not have, at all, the impression of somebody who would score minimally on those tests. I feel that it is not at all accurate" (participant with an ambulatory disability).

Michal Luria, Matt Scherer, Dhanaraj Thakur, Ariana Aboulafia, Henry Claypool, Wilneida Negrón

Another participant who scored in the fourth percentile on one of the gamified assessments shared a similar experience:

> *"I take depositions, I appear in court. I write and argue motions. I do so much more than what that four says about how stupid I am. I'm not stupid. [...] I really don't think that those tests were accurately measuring the actual abilities of anybody to understand complex information, decode it accurately and answer questions"* *(participant with a brain injury).*

But these issues go beyond individual tests or elements of test design. Many participants raised concerns that these tests would generally misrepresent their skills and abilities with negative implications for their and others' hiring. In fact, several concluded that based on their process and results, they do not believe they could get a job that makes use of these assessments – not even the job they hold now.

Participants felt that not only were the assessments improperly measuring what they intended to measure, they were also deficient in capturing the strengths that participants could offer to an employer, painting an incomplete picture of the assessment-taker. One interviewee elaborated:

> *"These scores don't really capture me. I don't think these scores could really capture anyone, regardless of neurotypical or neurodivergent. [If you had] everybody in the room, people who really know each other, and then they all take this test and look at the scores [...] Then everybody could see, "Oh, maybe the tests are wrong. Maybe we shouldn't use this to hire people. Maybe we're not getting the full picture."* *(participant with an ambulatory disability).*

2.3 Particular concerns around automated assessments of video interviews

The video interviewing system participants interacted with in this study leveraged AI to analyze applicants' facial expressions, and many participants were especially concerned that the automated analysis would discriminate against them based on their disability.

A participant with a cognitive disability, among others, explained:

> *"Judging facial expression is completely ableist to autistic people [...] judging them based on that could completely lock them out on something they can't control [...] [Employers] should actually watch the interview to see what their possible employees [are] like"* *(participant with a cognitive disability).*

Likewise, a participant with a vision disability pointed out that "*blind people might not know how to continually smile*," and that having a neutral expression for many should not be "*a failing of the test.*" Another participant expressed concern about how this kind of automatic analysis would process their physical disability:

> "*My disability involves a severe facial disfigurement from a shotgun blast. My face in resting position looks different. I can't really smile or frown. My mouth is just an opening in my face that surgeons created. I would say it's certainly not going to be a one-size-fits-all all tool.*"

The automatic video analysis supports participants' arguments that it may disadvantage disabled workers. The video analysis used in this study used AI to classify facial expressions in each recorded response as either "happy," "neutral," "disgust," "angry," "sad," or "surprise." Of the 14 participants who were able to complete the video interview assessment, which included recording nine different responses, only one participant's facial expressions were ever classified as "happy," and only in two out of their nine interview answers. For this participant, their facial expressions were classified as "disgust" in the seven other answers. The overwhelming majority of facial expression classifications were "neutral" (40% of all classifications across participants who completed the video interview assessment), "angry" (18%), or "surprise" (17%). That said, we do not have information regarding the distribution of facial expressions among non-disabled respondents.

2.4 Remote and asynchronous elements may benefit some, but otherwise digitized assessments offer few benefits to disabled workers

While participants generally concluded that assessments disadvantaged disabled workers, some noted that the ability to engage in a hiring process remotely and asynchronously could be beneficial:

> "*One can be comfortable doing [tests] from one's home, and it might relieve some psychological pressures*" (participant with a vision disability).

One participant pointed out that asynchronous assessments can be especially valuable to "*people who are already employed, perhaps in shift work where they don't have consistent time during working hours to be able to do an interview*" (participant with cognitive and ambulatory disabilities).

Some participants also noted that the format of the recorded, asynchronous interview, in which they had time to prepare an answer to a given question prior to recording it, also allowed for some flexibility and accommodated some disabilities:

> *"You have time to prepare and the employer has time to watch [the video responses] at the employer's leisure"* (participant with a vision disability).

But not all participants preferred this asynchronous setup. Some found asynchronous video interviews difficult. As one participant put it: "*I was responding to the questions, but I felt like I was talking to myself, so I felt like I was just BS-ing my way through.*" (participant with vision, hearing, ambulatory and cognitive disabilities).

Similarly, a participant with an ambulatory disability described:

> *"When I interview, and I answer these questions, I try to match the level of engagement of my interviewer. I wear my emotions on my face. I think the difficulty here was that I wasn't talking to a person. I was just talking [...] it was just hard not having somebody else to talk to and not knowing how my ideas and my thoughts and stories were being received."*

3. Digitized Assessments May Force Applicants to Reveal Disabilities Unwillingly

Federal law makes it illegal for employers to inquire into a prospective employee's disability status or request a medical examination until the employer makes a conditional job offer.[7] Yet most participants agreed that it was very hard, if not impossible, to go through the assessments without revealing one's disability status to the employer – people's disability could be inferred from their performance or from a request for accommodations. Several participants thought that in many instances, the only alternative to revealing their disability would be to "*bomb that part of the test*" (participant with a vision disability). Another referred specifically to the emotional intelligence test, that asked participants to identify an emotion portrayed in an image, as an example of a disability-revealing assessment, given some participants' inability to see the image clearly, let along the emotion it is portraying:

---

7   *See* U.S. Equal Employment Opportunity Commission, Enforcement Guidance on Disability-Related Inquiries and Medical Examinations of Employees under the ADA, July 26, 2000, https://www.eeoc.gov/laws/guidance/enforcement-guidance-disability-related-inquiries-and-medical-examinations-employees.

"[This assessment would] force [job seekers] to out themselves to this potential employer as a person with a disability whether or not they were ready to do that" (participant with a vision disability).

A handful of participants did note that if a digitized assessment reveals their disability, that may simply accelerate conversations with the employer about their disability, which would eventually occur anyway:

"I think there is a debate among blind professionals and blind attorneys in particular as to when and how do you disclose disability. I've always been of the mindset that I'll disclose it as early as I can because if it's going to be a problem for you [the employer], I don't want it to be a problem for me [...] that's not a person I want to work for" (participant with a vision disability).

Nevertheless, most participants did not view the forced revealing of their disability as a positive outcome. Moreover, at least one participant felt that the risk of revealing their disability created a chilling effect on how they completed the assessments – instead of focusing on the task, they described spending most of their cognitive effort in making sure they did not unintentionally reveal their disability through their performance or due to the answers they gave.

"It honestly just makes you feel bad without explaining how you can fix yourself or why [the assessment is] important. It doesn't seem important, especially if they give these tests for maybe minimum wage jobs. Having worked in retail, it's pointless and just frustrating."

4. Participants Found Digital Assessments Cognitively and Emotionally Taxing

Throughout the study, participants' experiences showed again and again that digitized assessments can be challenging, frustrating, demanding, and overwhelming for people with disabilities. In many cases, this meant participants were unable to complete the tests or underperformed when doing so. Even though a study team member walked participants through taking the tests, as many as nine of 17 participants in our study (53%) were unable to complete at least one of the assessments due to their disability. Seven more (41%) scored less than the fifth percentile on at least one of the tests. For example, the test that participants collectively performed worst on was the gamified arithmetic test; nine participants got an 'E' letter grade (scoring lower than the fifth percentile), and seven participants were unable to complete it at all — a total of 16 out of the 17 participants in the study were either unable to complete this test or scored lower than the fifth percentile.

In addition to the specific accessibility hurdles previously mentioned in, participants' experiences of attempting assessments were emotionally and mentally burdensome.

Michal Luria, Matt Scherer, Dhanaraj Thakur, Ariana Aboulafia, Henry Claypool, Wilneida Negrón

"You might get defeated in some of the questions and that might fall into like, 'Well, maybe I shouldn't work because I can't even answer these little simple questions.'"

One participant shared:

"Overall I felt frustrated, sh**ty, wasting time and very stressed. I ended up feeling overwhelmed by those things because obviously, they appear to some people as pretty easy, but they did not seem that way to me at all" (participant with an ambulatory disability).

Another participant similarly noted:

"It honestly just makes you feel bad without explaining how you can fix yourself or why [the assessment is] important. It doesn't seem important, especially if they give these tests for maybe minimum wage jobs. Having worked in retail, it's pointless and just frustrating" (participant with vision, cognitive, and other physical disabilities).

According to participants, this frustration partially stems from the many ways in which the assessments were inaccessible, creating a cumulative effect for people with disabilities. As a result, participants felt overwhelmed and under significant cognitive load:

"A lot of developers don't realize that when you have a person with multiple disabilities and [a] general nervous condition, switching attention is a big thing" (participant with an ambulatory disability).

Another participant with an ambulatory disability said they got so overwhelmed that they just "*started guessing and clicking stuff.*" A participant with low-vision shared how they spent all their cognitive capacity reading captions, instead of answering the questions at hand:

"It broke my brain. I was using so much energy trying to read the captions. I've got such limited vision that it was really challenging to read the captioning on [the assessments]. It gave me a bit of whiplash, honestly" (participant with a vision disability).

Not only were assessments emotionally and cognitively taxing, they also had potentially lasting impacts on participants' self-esteem and confidence in their ability to join the workforce:

"You might get defeated in some of the questions and that might fall into like, 'Well, maybe I shouldn't work because I can't even answer these little simple questions'" (participant with vision and Multiple Sclerosis-related disabilities).

"If it were the real world, I would feel inadequate and insufficient because I couldn't complete [the assessments], and that might psychologically affect an interviewee going forward with other exams or assessments, or even the actual interview for that position" (participant with a vision disability).

The immediate consequence of these cognitive and emotional costs, along with their lasting impact, can be a higher dropout rate from the hiring process among people with disabilities, who already have low employment rates compared to non-disabled people (Bureau of Labor Statistics, 2024):

"Honestly, if I was applying to a job, and they said, 'Here, take all these focus tests,' I'd just go try to find a different job [...] I wouldn't put myself through that" (participant with an ambulatory disability).

5. Participants Wanted Digitized Assessments to Include a Human in the Loop, be Supplementary, and be Transparent

Participants in our study varied on their overarching perspectives on digitized assessments. Some argued that any use of the studied assessments, even ones that address the current accessibility gaps, would discriminate against people with disabilities. Others were more supportive of creating accessible versions. Whether in one group or the other, participants agreed on several boundaries that should not be crossed.

5.1 Human in the loop

The most common constraint participants suggested was making sure that a human was in the loop throughout any use of digitized assessments. While the term "human in the loop" can convey many different meanings, in this context participants agreed that having the employer solely rely on outputs of digitized assessments or other automated tools to evaluate and make decisions about candidates would be very concerning. According to one participant, employers "*should not ever use [digitized assessments] unless they're going to meet with the individual employee to speak about their results*" (participant with cognitive, vision, self-care, and independent living disabilities).

Michal Luria, Matt Scherer, Dhanaraj Thakur, Ariana Aboulafia, Henry Claypool, Wilneida Negrón

Participants also expressed that certain automated functionalities of the assessments were fundamentally ableist and should be replaced by human review rather than considered alongside human evaluations. Instead of relying on automated scoring interfaces that digitized assessments allow, employers should to be active in the hiring process in a way that allows them to make key judgments themselves, especially with diverse and marginalized groups:

> *"If you're going to assess a candidate as their whole self across marginalization and various factors of diversity, you have to actually have face-to-face contact with them to be able to assess them accurately" (participant with cognitive and ambulatory disabilities).*

5.2 Supplementing, not screening

Participants consistently noted that these digitized assessments should not be used as a way to screen candidates out, but only for providing supplementary information about them, because, while the tests may assess "*the skill set of an existing candidate*," it cannot tell "*much about the person themselves*" (participant with a vision disability). Thus, if they are used at all, the digitized assessments should be used "*with a grain of salt – they should be used in conjunction with consultation of the résumé, and in conjunction with actual talking with the person*" (participant with an ambulatory disability). As highlighted earlier, employers should also not use fundamentally discriminatory aspects of digitized assessments, such as automated facial analysis.

Several participants speculated that instead of screening, digitized assessments could be a way to determine which role a candidate would be best suited for among several possible positions. According to a participant with ambulatory disability, employers might think, "*Oh, this person doesn't do as well with risk-taking, so we're going to move them into a position where risk-taking is a small percentage of their job.*"

Participants had a range of perspectives on when digitized assessments would be most suitable to integrate within the broader hiring process and human review. Many participants thought it would make most sense to do the tests first, and then have an interview with the employer and discuss the tests and their outcomes as needed:

> *"If you do [the tests] before the interview, you already have a feel for the person that you're interviewing, you can compare the test scores to the actual interview" (participant with an ambulatory disability).*

Other participants thought that "*if [tests] were presented as the first component of a multi-step hiring process*" then as job seekers, they would be "*discouraged from continuing going through the tests*" (participant with a vision disability) given how difficult they were for people with disabilities to complete. Another participant similarly pointed out that it "*might psychologically affect an interviewee going forward with other exams or assessments, or even the actual interview for that position*" (participant with a vision disability) if they were to go through tests as a first stage.

Nevertheless, both approaches shared the assumption that hiring assessments would not be a screening step in the process, but be combined with direct communication and conversation between employer and employee.

5.3 Transparency throughout

A participant … argued that "it's a little shady" to "make someone take a test that they don't know what it's assessing […] especially when there's such a huge power differential between an employer and a potential employee."

Participants agreed that "*it would be better to be more transparent*" (participant with a hearing disability) across all uses of assessments in the hiring process, including when they will be used, what they are being used for, and how they may impact the hiring process: "*In a perfect world,*" said a participant with hearing, vision and cognitive disabilities, "*100% transparency. I get that that's not a typical reality, but I'll say – as much as possible.*" Another participant described how more transparency about hiring technology can support job seekers:

> "*If I know what my employer's looking for while I'm taking the test, then I can apply that and take the tests with the mentality of what they're looking for, which can reduce the stress and also help me focus on my answers […] especially from my experience as a person that's disabled, I'd want as much information to make logic-based decisions, and to understand what they're looking for [from] me while I'm doing this so that I don't focus on the wrong thing. For example, if they want you to be as fast as possible, they should explain that, or if they want you to be as in-depth as possible, they should explain that*" (participant with cognitive and multiple physical disabilities).

A participant with cognitive and ambulatory disabilities argued that "*it's a little shady*" to "*make someone take a test that they don't know what it's assessing […] especially when there's such a huge power differential between an employer and a potential employee.*" They suggested that the best solution would be to offer "*full transparency on the job posting, letting people know, 'We are going to use these throughout the employment process,' and by applying you are consenting to [the] use of those.*"

Michal Luria, Matt Scherer, Dhanaraj Thakur, Ariana Aboulafia, Henry Claypool, Wilneida Negrón

Transparency was also important to several participants who were against the use of digitized assessments, and noted that transparency about the use of assessments in a particular position would allow them to make a decision to "*avoid applying for that job*" (participant with an ambulatory disability).

Discussion and Implications
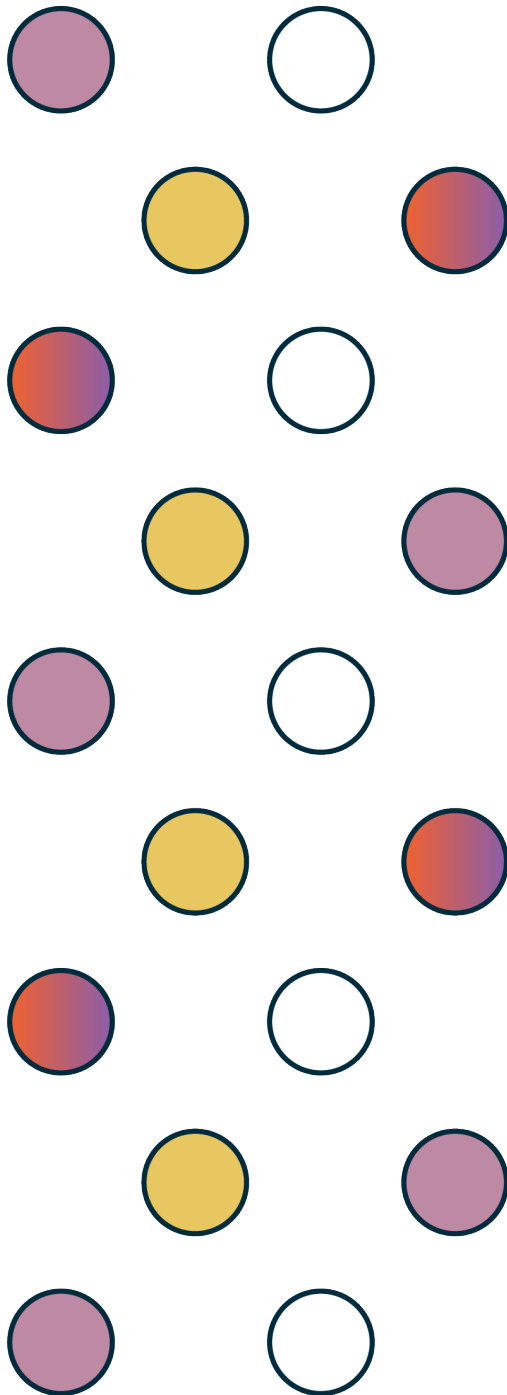
During the study, participants with disabilities were prevented from effectively completing digitized assessments and felt disadvantaged in two related ways. The first was in the process itself — completing the assessments was time-consuming, frustrating, and emotionally taxing as participants with many different types of disabilities repeatedly encountered accessibility barriers. In some instances, these challenges caused participants to drop out from the study or prevented them from completing some of the assessments. Even among those participants who completed the study, several stated during their interviews that if they had been asked to complete these assessments as part of a real-world hiring process, they likely would have dropped out before completing the process.

The second disadvantage to disabled participants played out in their assessment results and the potential impact on hiring outcomes in real hiring situations; they believed the digitized assessments were discriminatory in nature, and thus hindered an already challenging process for disabled people to apply to a job. The scores participants received not only caused more stress, discomfort, and cognitive effort in the next assessments, but participants felt their often low scores were unjustified and discriminatory, misrepresenting their ability to perform a job or join the workforce as a whole. Participants' experiences underscored that, for several reasons, digitized assessments do not work as intended for people with disabilities, and as we outlined throughout the findings — consistently integrate and systematize ableist assumptions.

In some cases, participants thought that aspects of assessments could be improved with various adjustments, like including a range of tools, accommodations, and alternative options for diverse disabilities, from simple accommodations like alt-text to more elaborate options like audio recordings for assessment questions. According to participants, introducing more accommodations would help level the playing field for candidates and would make the hiring process more fair.

Indeed, some of the barriers in the assessments were in part due to their design not conforming with longstanding web accessibility guidelines.[8] A more accessible design may result in different experiences for some

8   Web Content Accessibility Guidelines (WCAG): https://www.w3.org/TR/WCAG21/

individuals. To further understand to what extent proper accessibility and the lack of adherence to guidelines plays in the overall disadvantage to people with disabilities, researchers may need to replicate our assessment in more accessible survey platforms or to redesign some of the questions to be more accessible.

That said, many participants thought that some assessments or aspects of the assessments were fundamentally discriminatory and could not be fixed with better accommodation, and as a result, advised against their use. In those cases, participants argued that nothing can be done to make the digitized assessments inclusive of people with disabilities. For instance, participants may score lower on a test as a result of a disability, even though the assessment does not measure an essential function of the job. Using an assessment under those circumstances may result in illegal discrimination.

Further, the expression of a disability could greatly vary between individuals, even with the same disability, and would require personalized adjustment, which cannot be automated. Thus, no matter how thoughtfully accommodations may be introduced, digitized assessments would still inherently discriminate against people with disabilities.

Due to these challenges, most participants agreed that if digitized assessments were still to be used, an acceptable way of using them would be to supplement in-person interaction, as opposed to a standalone step that aims to either rank or fully screen out candidates. In addition, transparency early and often in the process would better allow the individual job seeker to understand why and how digitized assessments are used in jobs they are applying to. Having that information can help job seekers adjust their focus and effort into what is being tested.

Recommendations

The results of this study suggest fundamental issues associated with using digitized assessments to evaluate disabled workers. This section presents some recommendations for developers and deployers of digitized assessments to help address these concerns. Because the goal of this report was to center the experiences and observations of disabled workers, the recommendations here are based on the comments and suggestions of study participants, as conveyed to the research team during the course of the project. They thus flow directly from the findings discussed above.

Concerns raised by participants might not be easily resolved, and some barriers may, in fact, be intractable. Consequently, these recommendations should be interpreted as suggestions for harm reduction rather than harm elimination.

1. Steps to Ensure the Assessments' Effectiveness and Accessibility

The first step question developers and deployers should ask is whether using the assessment is necessary, with no better alternatives to assess a particular skill set. If there is no better alternative, developers and deployers should then ask if the assessment is effective and fair, given their goal. Our findings suggest that developers often design digitized assessments without giving due consideration to the unique attributes and needs of disabled workers.

The second question should be whether digitized assessments are accessible, adhering to existing guidelines and best practices for accessibility and inclusive design. There is no shortage of published standards, guidelines, and best practices for digital accessibility. For example, the Web Content Accessibility Guidelines (WCAG)[9] were first introduced in 1999 and continue to be updated periodically (Caldwell et al., 2008), and the Partnership on Employment and Accessible Technologies (PEAT)[10] has a number of active initiatives aimed at ensuring hiring and workplace technologies are accessible to disabled workers.

---

9   Web Content Accessibility Guidelines (WCAG): https://www.w3.org/TR/WCAG21/

10   Partnership on Employment and Accessible Technologies (PEAT): https://www.peatworks.org/
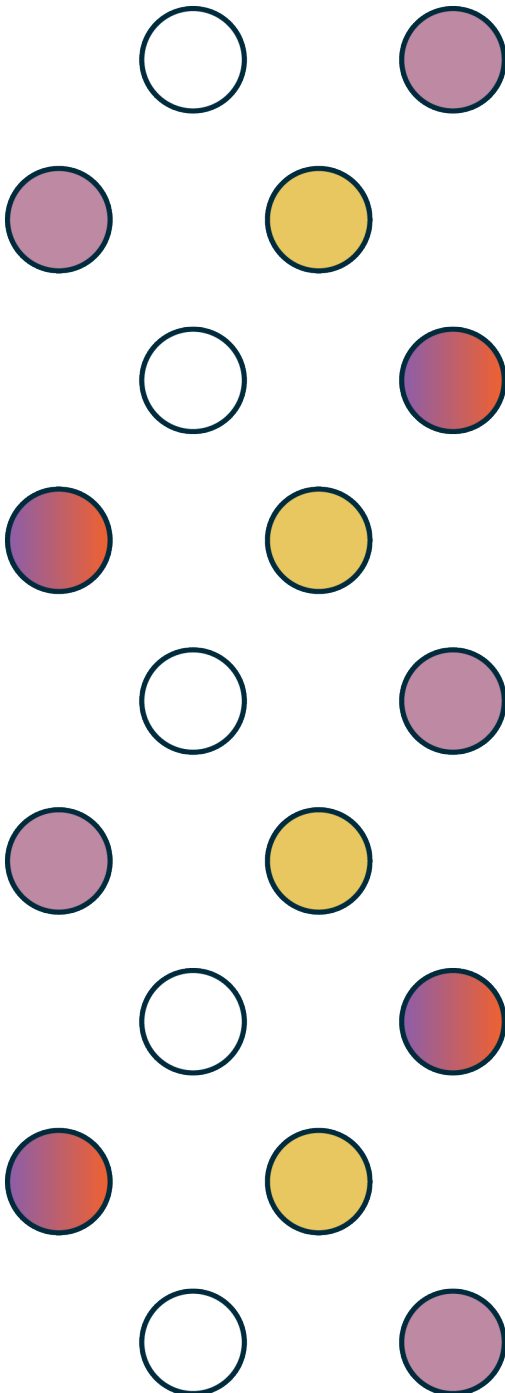
Employers can test whether a digitized assessment is both accessible and an effective way to measure the relevant knowledge and skills for a particular job by having workers with a wide range of disabilities test the assessment and provide feedback on aspects of the assessment or its implementation that may present accessibility barriers or prevent it from accurately measuring the skills and abilities of disabled workers.

If the discussed accessibility guidelines cannot be incorporated, or if a hiring tool's effectiveness and accessibility cannot be confirmed, the technology is likely to disadvantage disabled workers and thus should also not be used.

2. Steps to Mitigate Potential Biases During the Deployment of Digitized Assessments

In a case where an assessment is effective, fair and accessible, employers can take steps to further reduce biases and the potential of unfairly disadvantaging or screening out disabled workers:

1. **Keep humans in the loop.** Automated steps should be kept to a minimum, and employers should maximize human review at every crossroad to ensure assessments are attuned to individuals instead of assuming a one-size-fits-all assessment. A system should be established so that participants have adequate time to request accommodation — and a human is empowered to promptly act on such requests — prior to assessment. Moreover, because humans are biased to believe algorithmic outputs and under-correct erroneous outputs (Sele & Chugunova, 2024; Agudo et al., 2024), employers should adopt policies that specify the purpose of the hiring tool, the appropriate interpretation of its output, limitations on the system, and the weight to be given to its output (Quay-de la Vallee & Duarte, 2019).

2. **Use to supplement candidate information, not to screen out.** Digitized assessments and associated data should be used in combination with other information about candidates to make a holistic assessment of their suitability for a particular job.

3. **Be transparent about assessments' use and role.** Ensure disabled candidates, and all candidates, have enough information about when and how digitized assessments and other modern hiring technologies will be used. This will allow them to adequately request accommodation and have confidence in the fairness of the hiring process.

Appendix 1:
Study Participant Information

PID	Age	Worker Type	Disability
01	28	Attorney / Law Student	Cognitive, Ambulatory
02	23	Attorney / Law Student	Ambulatory
03	37	Attorney / Law Student	Vision
04	44	Attorney / Law Student	Vision
05	28	Attorney / Law Student	Vision
06	59	Attorney / Law Student	Brain Injury
07	29	Hourly Worker	Ambulatory, Self-care, Ind. Living
08	37	Hourly Worker	Hearing, Vision, Cognitive, Ambulatory, Ind. Living, Other: Autism
09	18	Hourly Worker	Cognitive
10	20	Hourly Worker	Autism
11	21	Hourly Worker	Hearing-Deaf
12	23	Hourly Worker	Ambulatory, Cognitive, Ind Living
13	25	Hourly Worker	Cognitive
14	34	Hourly Worker	Ambulatory, Ind Living
15	22	Hourly Worker	Cognitive, Independent living, Chronic pain (Ehlers-Danlos), Epilepsy, Narcolepsy, Vision
16	42	Hourly Worker	Vision, Multiple Sclerosis (MS)
17	51	Hourly Worker	Hearing, Vision, Cognitive, Independent living

▲   **Table 1. All participants in the study, including their reported disability, age, and their worker type.** Participants selected a disability from a set of predefined disabilities developed by the research team.

Other Demographic Information

Age

- **Range:** 19 to 59
- **Mean:** 31.2
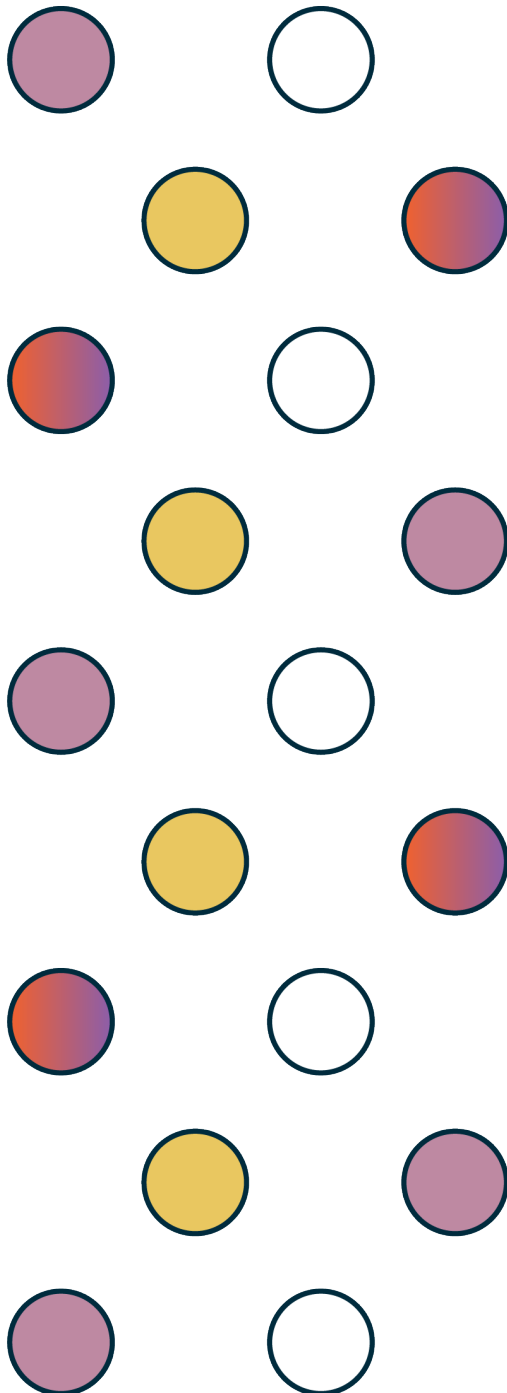- **Median:** 28

Gender

- **Women:** 8
- **Men:** 5
- **Non-binary:** 2
- **Genderqueer:** 1
- **Woman/non-binary:** 1

Race/ethnicity

- **White:** 8
- **Black:** 3
- **Hispanic or Latino:** 1
- **Pacific Islander:** 1
- **More than one race:** 1
- **Middle Eastern (written in):** 1

Appendix 2: Digitized Assessments Used

This is a description of the digitized assessments that participants were asked to take part in the study. We selected up to eight assessments for participants to complete—hourly workers were asked to complete seven tests, and lawyers were asked to complete eight tests. The tests were divided into two sittings to reduce cognitive load and the intensity for each, each lasting up to two hours. The assessments were selected based on a review of the marketing materials for hiring technology vendors and employment test preparation companies.

Work Personality Questionnaire (7-12 minutes long)

This test is a self-report assessment in which participants are asked to either agree or disagree with a set of work-related behavioral statements, such as "I enjoy working with others". The test is based on the Big Five model of personality. It includes 90 questions that attempt to gauge an individual's professional preferences and personality characteristics compared to other individuals.

Balloon Test Gamified Assessment (4-8 minutes long)

The test presents an illustrated deflated balloon on an illustrated forest background. Participants are asked to click a button to "inflate" each of 15 balloons and "collect" them before they burst. Each balloon bursts after a random number of inflations. A participant receives no tokens if the balloon bursts. The test claims to assess three aspects: (1) emotionality and risk taking, (2) work management and decision-s making and (3) sociability. This test simulates an assessment offered by Pymetrics, an AI company that offers a suite of gamified assessment. In 2022, Pymetrics was purchased by Harver, another vendor of automated employment tests.

Directional Attention Gamified Assessment
(7-13 minutes long)

Participants are presented with a row of seven illustrated objects that each face either right or left. Participants are asked to indicate which direction the center object is facing in for a series of 85 fast-paced sequences. The test sets out to evaluate cognitive attention ability, including selective attention and executive control. This test also simulates another Pymetrics gamified assessment.

Emotional Intelligence Gamified Assessment
(3-4 minutes long)

This assessment presents a series of 40 photographs of people expressing a range of emotions. Participants are asked to recognize and indicate which single emotion is being portrayed in each image (out of a list of eight set emotions), over a duration of three minutes. The goal of the test is to identify candidates' ability to recognize emotion and to evaluate some of their emotional intelligence. This assessment is similar to the Pictures of Facial Affect test developed by psychologist Paul Ekman and offered by Harver and a number of other employment test vendors.

Arithmetic Gamified Assessment
(5-6.5 minutes long)

In this test, participants are asked to solve a set of simple math problems. At every given moment, the screen shows four "bubbles" with math equations in different areas of the screen. On the top of the screen, there is a stated "target" number. Participants are asked to "pop" any bubbles that match the target number. For example, if the target number is 7, participants might "pop" 3+4 or 7x1. The test is time-constrained, and bubbles disappear and reappear with new equations every several seconds. It attempts to assess individuals' proficiency in quick arithmetic calculations.

Watson-Glaser, *lawyers only*
(25-40 minutes long)

The Watson-Glaser test aims to assess one's ability to make inferences and assumptions through logical reasoning in five categories: argument analysis, assumptions, deductions, inferences, and information interpretation. For example, for the

assumption category, participants were asked whether an assumption is made in the bolded part of a statement: "My friend should have less sugary drinks in his diet to better protect his teeth. Consuming too much sugar can be bad for your teeth." The test lasts a total of 40 minutes, during which 40 questions are presented.

Situational judgment test (25-60 minutes long)

This test presents a set of 18 real-life problem-solving scenarios situated in a workplace. For each, four courses of action are presented alongside the scenario. Respondents are asked to indicate which of the four would be the best and worst way to respond to the situation. For example: "You work as a Manager on the shop floor of a retail outlet when you are approached by a customer who is clearly upset. You are unable to make out exactly what the customer is saying, but you are mindful of the fact that the customer's tone is drawing the attention of other customers. What do you do?" According to the website, this assessment tests nine different work-related competencies: solving problems, customer-centered, delivering results, planning and organizing, leading, innovating, working with colleagues, communicating, and influencing and persuading.

Video interview (15-20 minutes long)

For the final test, participants were asked to do a simulated asynchronous job interview by recording videos of answers to a set of nine interview questions. The interview consisted of questions that were specifically directed towards retail job seekers (for hourly workers) or law job seekers (for lawyers). The interface presented one question at a time, and asked participants to begin recording an answer when they were ready. Each answer was limited to 2:30 minutes. Participants were told prior to the assessment that it uses AI to analyze their facial expressions. This particular assessment did not use AI or other automated tools to transcribe or analyze candidates' spoken responses, though many AI-powered video interview tools used by employers may include such functionality.

Michal Luria, Matt Scherer, Dhanaraj Thakur, Ariana Aboulafia, Henry Claypool, Wilneida Negrón

References

ACLU. (2024). *ACLU complaint to the FTC regarding Aon Consulting, Inc*. American Civil Liberties Union. https://www.aclu.org/documents/aclu-complaint-to-the-ftc-regarding-aon-consulting-inc [perma.cc/Q89V-B7NK]

Agudo, U., Liberal, K. G., Arrese, M., & Matute, H. (2024). The impact of AI errors in a human-in-the-loop process. *Cognitive Research: Principles and Implications, 9*(1), 1. https://doi.org/10.1186/s41235-023-00529-3 [perma.cc/LZ7A-UXBY]

Akselrod, O., & Venzke, C. (2023, August 23). How Artificial Intelligence Might Prevent You From Getting Hired | ACLU. *American Civil Liberties Union*. https://www.aclu.org/news/racial-justice/how-artificial-intelligence-might-prevent-you-from-getting-hired [perma.cc/7M3M-KYKQ]

Birhane, A. (2022). *Automating Ambiguity: Challenges and Pitfalls of Artificial Intelligence* (arXiv:2206.04179). arXiv. http://arxiv.org/abs/2206.04179 [perma.cc/7DQU-QHT2]

Bonaccio, S., Connelly, C. E., Gellatly, I. R., Jetha, A., & Martin Ginis, K. A. (2020). The Participation of People with Disabilities in the Workplace Across the Employment Cycle: Employer Concerns and Research Evidence. *Journal of Business and Psychology, 35*(2), 135–158. https://doi.org/10.1007/s10869-018-9602-5 [perma.cc/T3U3-TWFZ]

Brown, L. X. Z., Shetty, R., & Richardson, M. (2020). Algorithm-driven Hiring Tools: Innovative Recruitment or Expedited Disability Discrimination? *Center for Democracy and Technology*. https://cdt.org/insights/report-algorithm-driven-hiring-tools-innovative-recruitment-or-expedited-disability-discrimination/ [perma.cc/NV4P-UW5H]

Bureau of Labor Statistics. (2024). *Persons with a Disability: Labor Force Characteristics Summary - 2023 A01 Results*. Bureau of Labor Statistics. https://www.bls.gov/news.release/disabl.nr0.htm [perma.cc/HVW8-FNU4]

Caldwell, B., Cooper, M., Reid, L. G., Vanderheiden, G., Chisholm, W., Slatin, J., & White, J. (2008). Web content accessibility guidelines (WCAG) 2.0. *WWW Consortium (W3C), 290*(1–34), 5–12. https://travesia.mcu.es/server/api/core/bitstreams/6c7e00c8-cd2d-4066-bba0-2203ba2281ad/content [https://perma.cc/7G3K-FSPF]

Cottrell, J. M., Newman, D. A., & Roisman, G. I. (2015). Explaining the black–white gap in cognitive test scores: Toward a theory of adverse impact. *Journal of Applied Psychology, 100*(6), 1713–1736. https://doi.org/10.1037/apl0000020 [perma.cc/83X9-UMLL]

Dastin, J. (2018, October 11). Insight—Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*. https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG/ [https://perma.cc/M4MC-685S]

Fruchterman, J., & Mellea, J. (2018). *Expanding Employment Success for People with Disabilities*. Benetech. https://benetech.org/wp-content/uploads/2018/11/Tech-and-Disability-Employment-Report-November-2018.pdf [https://perma.cc/52SQ-6H5G]

Gershgorn, D. (2018, October 22). *Companies are on the hook if their hiring algorithms are biased*. Quartz. https://qz.com/1427621/companies-are-on-the-hook-if-their-hiring-algorithms-are-biased [perma.cc/U289-BE3P]

Glasser, A., Kushalnagar, K., & Kushalnagar, R. (2017). Deaf, Hard of Hearing, and Hearing Perspectives on using Automatic Speech Recognition in Conversation. *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*, 427–432. https://doi.org/10.1145/3132525.3134781 [perma.cc/Z54B-UW2W]

Glazko, K., Mohammed, Y., Kosa, B., Potluri, V., & Mankoff, J. (2024). Identifying and Improving Disability Bias in GPT-Based Resume Screening. *The 2024 ACM Conference on Fairness, Accountability, and Transparency*, 687–700. https://doi.org/10.1145/3630106.3658933 [perma.cc/64YY-BXSS]

Harper, R. (2024). *The Skinny on Accessibility Overlays – Colorado Virtual Library*. https://www.coloradovirtuallibrary.org/learning/edit/the-skinny-on-accessibility-overlays/ [perma.cc/G63C-3TN3]

Hidalgo Lopez, J. C., Sandeep, S., Wright, M., Wandell, G. M., & Law, A. B. (2023). Quantifying and Improving the Performance of Speech Recognition Systems on Dysphonic Speech. *Otolaryngology–Head and Neck Surgery, 168*(5), 1130–1138. https://aao-hnsfjournals.onlinelibrary.wiley.com/doi/abs/10.1002/ohn.170 [https://perma.cc/SV8P-YSUQ]

Hsu, A. (2023, January 31). Can bots discriminate? It's a big question as companies use AI for hiring. *NPR*. https://www.npr.org/2023/01/31/1152652093/ai-artificial-intelligence-bot-hiring-eeoc-discrimination [perma.cc/4WDE-KEE6]

Hunkenschroer, A. L., & Luetge, C. (2022). Ethics of AI-Enabled Recruiting and Selection: A Review and Research Agenda. *Journal of Business Ethics, 178*(4), 977–1007. https://doi.org/10.1007/s10551-022-05049-6 [perma.cc/MQG8-NVV7]

Kim, P., & Bodie, M. T. (2020). *Artificial Intelligence and the Challenges of Workplace Discrimination and Privacy*. ABAJ Lab. & Emp. L. 35 (2020): 289. https://heinonline.org/HOL/LandingPage?handle=hein.journals/lablaw35&div=23&id [https://perma.cc/WK6U-47E4]

Mimbela, R., & Akselrod, O. (2024, May 30). The Long History of Discrimination in Job Hiring Assessments. *American Civil Liberties Union*. https://www.aclu.org/news/racial-justice/the-long-history-of-discrimination-in-job-hiring-assessments [perma.cc/7Q3M-SA97]

National Trends in Disability Employment. (2024). *nTIDE July 2024 Jobs Report: People with Disabilities Hold Steady in Labor Market Despite Federal Reserve's Attempts to Slow Economy | Kessler Foundation*. https://kesslerfoundation.org/press-release/ntide-july-2024-jobs-report-people-disabilities-hold-steady-labor-market-despite [https://perma.cc/9Q2S-RY65]

Outtz, J. L., & Newman, D. A. (2009). A Theory of Adverse Impact. In *Adverse Impact*. (pp. 53-94). Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9780203848418-5/theory-adverse-impact-james-outtz-daniel-newman [https://perma.cc/9JUV-DWG8]

Papinchock, J. M., Rosenbaum, A. L., & Dunleavy, E. M. (2023). Potential Impact of Disabilities and Neurodiversity on the Constructs Measured by Selection Procedures. In T. Kantrowitz, D. H. Reynolds, & J. Scott (Eds.), *Talent Assessment: Embracing Innovation and Mitigating Risk in the Digital Age* (p. 0). Oxford University Press. https://doi.org/10.1093/oso/9780197611050.003.0016 [perma.cc/LW8H-MQJZ]

Michal Luria, Matt Scherer, Dhanaraj Thakur, Ariana Aboulafia, Henry Claypool, Wilneida Negrón

Persson, A. (2016). Implicit Bias in Predictive Data Profiling Within Recruitments. In A. Lehmann, D. Whitehouse, S. Fischer-Hübner, L. Fritsch, & C. Raab (Eds.), *Privacy and Identity Management*. (pp. 212–230). Springer International Publishing. https://doi.org/10.1007/978-3-319-55783-0_15 [perma.cc/55VN-YY8D]

Quay-de la Vallee, H., & Duarte, N. (2019). Algorithmic Systems in Education: Incorporating Equity and Fairness When Using Student Data. *Center for Democracy & Technology*. https://cdt.org/wp-content/uploads/2019/08/2019-08-08-Digital-Decision-making-Brief-FINAL.pdf [perma.cc/3CBX-KEL3]

Raghavan, M., Barocas, S., Kleinberg, J., & Levy, K. (2020). Mitigating Bias in Algorithmic Hiring: Evaluating Claims and Practices. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 469–481. https://doi.org/10.1145/3351095.3372828 [perma.cc/WW69-6YXA]

Rieke, A., & Bogen, M. (2018, December 10). *Help Wanted: An Examination of Hiring Algorithms, Equity, and Bias*. Upturn. https://upturn.org/work/help-wanted/ [https://perma.cc/CWM4-YJ59]

Sackett, P. R., Zhang, C., Berry, C. M., & Lievens, F. (2022). Revisiting meta-analytic estimates of validity in personnel selection: Addressing systematic overcorrection for restriction of range. *Journal of Applied Psychology, 107*(11), 2040–2068. https://doi.org/10.1037/apl0000994 [perma.cc/LG3F-84NF]

Savage, D. D., & Bales, R. (2016). Video Games in Job Interviews: Using Algorithms to Minimize Discrimination and Unconscious Bias. *ABA Journal of Labor and Employment Law, 32*, 211. https://heinonline.org/HOL/LandingPage?handle=hein.journals/lablaw32&div=18&id= [https://perma.cc/PRD4-M9CZ]

Sele, D., & Chugunova, M. (2024). Putting a human in the loop: Increasing uptake, but decreasing accuracy of automated decision-making. *PLOS ONE, 19*(2), e0298037. https://doi.org/10.1371/journal.pone.0298037 [perma.cc/FNP5-ARZF]

Stark, L., View Profile, Hoey, J., & View Profile. (2021). The Ethics of Emotion in Artificial Intelligence Systems. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 782–793. https://doi.org/10.1145/3442188.3445939 [https://perma.cc/A4LD-FQXL]

The Partnership on Employment & Accessible Technology (*PEAT*). (n.d.). Peatworks. Retrieved September 6, 2024, from https://www.peatworks.org/ [perma.cc/7Z66-UWSC]

Tu, M., Wisler, A., Berisha, V., & Liss, J. M. (2016). The relationship between perceptual disturbances in dysarthric speech and automatic speech recognition performance. *The Journal of the Acoustical Society of America, 140*(5), EL416. https://doi.org/10.1121/1.4967208 [perma.cc/5BYT-TFAT]

Williams, D. P. (2024). Disabling AI: Biases and Values Embedded in Artificial Intelligence. In D. J. Gunkel (Ed.), *Handbook on the Ethics of Artificial Intelligence* (pp. 246–261). Edward Elgar Publishing. https://www.elgaronline.com/edcollchap/book/9781803926728/book-part-9781803926728-22.xml [https://perma.cc/MXG6-FU6Z]

cdt.org

cdt.org/contact

Center for Democracy & Technology
1401 K Street NW, Suite 200
Washington, D.C. 20005

202-637-9800

@CenDemTech

@cendemtech.bsky.social

cdt CENTER FOR
DEMOCRACY
& TECHNOLOGY