

# Equitable Employment Solutions: Enhancing Skills Matching for Marginalized Groups

A Major Qualifying Project

Submitted to the faculty of

WORCESTER POLYTECHNIC INSTITUTE

In partial fulfillment of the requirements for the

Degree of Bachelor of Science

By:

Sasha Daraskevich

Ben Erwin

Mikaela Milch

Nolan Willoughby

Date:

March 1, 2024

Report Submitted to:

Professors Daniel Reichman and Andrew Trapp

Worcester Polytechnic Institute

Intrare, Sponsor

Mexico City, Mexico



# WPI

*This report represents the work of one or more WPI undergraduate students submitted to the faculty as evidence of completion of a degree requirement. WPI routinely publishes these reports on the web without editorial or peer review.*

# Abstract

Marginalized demographic groups face discrimination in the labor market, limiting job opportunities and further isolating them from society. Intrare, a social startup based in Mexico City, Mexico strives to empower vulnerable groups by connecting them with equitable employment opportunities. The goal of this Major Qualifying Project was to enhance Intrare's job matching algorithm with a soft skills assessment and accompanying interface. Following the engineering design process, our project defined requirements, selected relevant soft skills, developed an online questionnaire, and conducted user testing to refine the interface and assessment. The soft skills assessment is anticipated to provide a well-rounded view of candidates. This will strengthen Intrare's existing job matching algorithm to provide more accurate and sustainable job matches, allowing the users to feel confident in their abilities. The team recommends administering focus groups and monitoring trends in user responses to the questionnaire then adapting accordingly.

# Acknowledgements

We would like to thank everyone who helped make this project possible:

- Our project sponsor, Intrare, for providing us with the opportunity to work on this project. Specifically, thank you to Hannah Töpler, Carlos Turrent, Carlos Martinez, Alejandra Matheus, and Sofia Campa for attending bi-weekly meetings and guiding us throughout the duration of the project.
- Our project advisors, Dr. Andrew Trapp and Dr. Daniel Reichman for their dedication to the project team, their willingness to think outside of the box, and for leading us through the MQP process.
- WPI Data Science PhD candidate Marcela Vasconcellos for introducing us to skills assessments, providing us with her preliminary research, and answering our questions.
- Professor Jed Lindholm for generously combing through each proposed question, providing thoughtful feedback, and suggesting rephrasings.
- The WPI Office of Diversity, Inclusion, and Multicultural Education for evaluating our questionnaire to ensure we were being considerate of user's lived experiences.
- Professor Lane Harrison for providing expert feedback on our proposed data visualization, allowing us to produce a quality product.
- WPI Data Science PhD candidate Kathleen Cachel for taking the time to discuss fairness in algorithms with us and helping us determine what approach we should take.
- WPI Mechanical Engineering and Robotics Engineering Undergraduate Student Pau Acolea for generously translating the questionnaire introduction and questions into Spanish.

# Table of Authorship

<b>Section</b>	<b>Author</b>	<b>Editor</b>
Chapter 1. Introduction	All	All
Chapter 2. Background	All	All
2.1: Intrare: Job Matching for Vulnerable Groups	Sasha	All
2.2: Context	Nolan	Sasha
2.3: Skills Matching	Ben	Nolan
2.4: A Review of Prominent Skills Assessment	Mikaela	Ben
2.5: Implementation of Fairness in Algorithms	Sasha	Mikaela
Chapter 3. Design	All	All
3.1: Defining the Problem Statement	All	All
3.2: Designing the Assessment	All	All
3.3: Evaluating the Assessment and Interface	Mikaela	All
Chapter 4. Methodology	All	All
4.1: Questions	All	All
4.2: Front End	Mikaela	Ben
4.3: Back End	Mikaela	Sasha
4.4: Skills Vector	Sasha	Mikaela
Chapter 5. Results	All	All
5.1: Front End Results	Mikaela	Ben
5.2: Heuristics	Ben	Sasha
5.3: A/B Testing	Sasha	Nolan
Chapter 6. Recommendations & Conclusions	Nolan	Ben
Chapter 7. Reflection	All	All

# Table of Contents

<b>Abstract</b> .....	i
<b>Acknowledgements</b> .....	ii
<b>Table of Authorship</b> .....	iii
<b>List of Figures</b> .....	viii
<b>List of Tables</b> .....	x
<b>1. Introduction</b> .....	1
<b>2. Background</b> .....	5
2.1 Intraire: Job Matching for Vulnerable Groups.....	5
2.2 Context .....	6
2.2.1 Vulnerable Groups in Society .....	6
2.2.2 Job Matching, Data, and Analytics with Vulnerable Groups .....	7
2.2.3 Recommended Mitigation Techniques .....	8
2.3 Skills Matching .....	9
2.3.1 Skills Matching Overview .....	9
2.3.2 Hard versus Soft Skills.....	10
2.3.3 Analysis of Benefits and Drawbacks .....	11
2.4 A Review of Prominent Skills Assessments.....	11
2.4.1 Pymetrics.....	12
2.4.2 Koru .....	13
2.4.3 NADINE .....	14
2.5 Implementation of Fairness in Algorithms .....	15
2.5.1 Overview of Fairness .....	15
2.5.2 Applications within Algorithms.....	17
2.5.3 Algorithmic Fairness in Action: Skills Matching .....	17

2.5.4 Further considerations.....	18
<b>3. Design</b> .....	19
3.1 Defining the Problem Statement.....	19
3.1.1 Problem: Intrare.....	19
3.1.2 Project Requirements.....	20
3.1.3 Constraints.....	20
3.1.4 Project Goals.....	21
3.2 Designing the Assessment.....	21
3.2.1 Creating a Comprehensive Skill Set.....	22
3.2.2 Limiting Bias in Pre-Employment Assessment Type.....	24
3.2.3 Develop Questions for Assessment.....	25
3.2.4 Scoring of Assessment.....	29
3.2.5 Designing an Accessible Platform for the Assessment.....	30
3.3 Evaluating the Assessment and Interface.....	33
3.3.1 Interface Testing.....	33
3.3.2 Assessment testing.....	35
<b>4. Methodology</b> .....	37
4.1 Questions.....	37
4.1.1 Original Draft Research.....	37
4.1.2 Simplification and Review.....	41
4.1.3 Milestones.....	42
4.1.4 Order.....	43
4.1.5 Translation.....	43
4.2 Front End.....	44
4.1.1 Prototyping.....	44

4.1.2 Creating a Front End with ReactJS .....	45
4.3 Back End .....	47
4.4 Skills Scoring Visualization.....	48
4.5 User Testing .....	51
4.5.2 Constraints .....	53
<b>5. Results .....</b>	<b>54</b>
5.1 Front End Results.....	54
5.1.1 Introduction.....	54
5.1.2 Questions.....	54
5.1.3 Conclusion Page.....	55
5.2 Heuristics .....	58
5.2.1 Protocol.....	58
5.2.2 Results.....	58
5.3 Preliminary User Results .....	60
5.3.1 Protocol .....	60
5.3.2 Results.....	60
<b>6. Recommendations &amp; Conclusion.....</b>	<b>65</b>
6.1 Focus Group.....	65
6.2 Observe and Adapt to Trends in Data.....	67
<b>7. Reflection .....</b>	<b>69</b>
7.1 Project Evolution .....	69
7.2 Constraints .....	70
7.3 Learning Outcomes .....	71
7.4 Working as a Team .....	72
<b>8. References .....</b>	<b>73</b>

<b>9. Appendix</b> .....	82
Appendix A: Heuristic Protocol with Intrare Team Members.....	82
Appendix B: A/B Testing Protocol.....	87
Appendix C: Focus Group with Intrare Users Protocol.....	88
Appendix D: App Translations .....	91
Appendix E: Encouragements.....	95



# List of Figures

Figure 1. Engineering Design Process ( <i>The 7 Steps of the Engineering Design Process</i> , 2023) ...	3
Figure 2. Balloon Analogue Risk Test (Lejuez et al., 2002) .....	12
Figure 3. Alternative Skills Vector (CodeSignal, 2023).....	15
Figure 4. Model for Survey Development (Strachota et al., 2013).....	26
Figure 5. First Draft of Wireframe Assessment.....	32
Figure 6. First Draft of Wireframe on Mobile Phone .....	32
Figure 7. First Draft of Wireframe Home Page .....	32
Figure 8. Milestones of Questionnaire.....	43
Figure 9. Wireframe Introduction Page.....	45
Figure 10. Wireframe Question Page.....	45
Figure 11. Circles Representing User Progress .....	46
Figure 12. Pop-up Encouragement .....	46
Figure 13. Database Schema.....	47
Figure 14. Iterative Draft of the Scoring Vector.....	50
Figure 15. Scoring Prototype .....	51
Figure 16. Scoring Prototype with Sub-skills .....	51
Figure 17. Language Barrier B Version.....	53
Figure 18. Font Size B Version.....	53
Figure 19. Questionnaire Introduction Page .....	56
Figure 20. Questionnaire Question Page .....	56
Figure 21. Question Page with Encouragement.....	57

Figure 22. Questionnaire Conclusion Page.....	57
Figure 23. Assessment Demographics by Gender .....	61
Figure 24. User Time Breakdown on Assessment.....	62
Figure 25. User Testing Demographics .....	63
Figure 26. Assessment Impact on Job Results.....	64

# List of Tables

Table 1. Intrare Skills Frequency Chart.....	22
Table 2. NADINE Soft Skills .....	23
Table 3. Top 10 Workplace Soft Skills.....	24
Table 4. Heuristics and Descriptions to Be Used in the Heuristic Evaluation .....	34
Table 5. Questions for the User Questionnaire.....	36
Table 6. Partial Heuristics Results .....	59

# 1. Introduction

Demographic groups such as racial minorities, women, and refugees face documented discrimination and inequity when participating in the labor market (Ashenfelter & Rees, 1973; Jones et al., 2021). These groups tend to make lower wages than their counterparts, and face greater risk of exploitation. An important aspect of achieving equity is ensuring people from marginalized demographic groups are able to find equitable and non-exploitative employment opportunities. One way of achieving this is through job matching. Because conventional methods of data collection and utilization are catered to majority demographic groups, these methods and uses can serve to further marginalize vulnerable populations (Ahani et al., 2021). This means that in order to employ analytical methods in the context of marginalized groups, adaptations must be made (Ruberg & Ruelos, 2020). One such analytical method is that of skills matching, which aims to consider candidate skills to predict job effectiveness. By using skills matching assessments, employers are able to effectively measure a candidate's soft and hard skills and potential for success in a given role (Joost et al., 2023). For example, prominent skill assessment vendors such as Pymetrics and Koru use neurological games to assess candidates, comparing their scores to those of successful employees ("Pymetrics Internal Demo Day Pitch", 2017; *Predictive Hiring*, 2018). These assessments may introduce bias by relying on subjective assessments and ranking candidates (Bogen & Rieke, 2018). The proposed NADINE assessment attempts to limit hiring discrimination against displaced peoples in Europe as well as recommends job categories that align with their abilities (Leligou et al., 2021). The recent rise of the power of analytics and machine learning has been implemented in varied applications and domains including the hiring process (Finocchiaro et al., 2021; Raghavan et al., 2020). While useful, this has been met with discussions

on what fairness means and how it can be correctly implemented to prevent societal harm and ensure beneficial outcomes.

One organization looking to implement these practices and empower members of vulnerable groups to find sustainable non-exploitative employment is Intrare, a Mexico City-based social startup founded in 2018. These employment opportunities assist individuals in becoming more financially independent and feeling more integrated in society. Intrare currently uses a job matching algorithm designed at WPI by Data Science PhD candidate Marcela Vasconcellos. This algorithm is based on personal information including prior experience and work preferences. However, the company is looking to strengthen the algorithm through the design and implementation of a pre-employment assessment to conduct skills matching.

The goal of our project was to enhance Intrare's job matching algorithm with a soft skills assessment and corresponding interface. To accomplish this, we followed the engineering design process (Figure 1) by first defining the problem, the requirements, the constraints, and ultimate goal. Then, based on our literature review, we brainstormed and determined how to best develop our assessment. First, we selected the soft skills to assess. The set of soft skills was based on current Intrare job descriptions, skills evaluated by existing skills assessments, skills commonly identified in academic research, and working with Intrare staff who are experts in the field. Next, we evaluated existing pre-employment assessment within the context of the project, current practices, their validity, and potential to create underlying biases. From there, we developed an online questionnaire to elicit the selected skills and assign scores to feed into an existing many-to-many matching algorithm. To implement the questionnaire, we developed a wireframe with accessibility as a primary focus, and used the wireframe design to build the assessment's user interface. With the questionnaire developed and a scoring system set, we then developed a scoring

vector to present users with their scores, which could then be shared with employers and used for personal development. This scoring vector was then added to the user interface. Then, after developing the assessment and interface, we conducted user testing to determine what adjustments to the original questionnaire and prototype were necessary. We also used a heuristic evaluation with Intrare employees because it is ideal when working with a small group of subject matter experts (Nielsen & Molich, 1990). We used the results from this testing and feedback from the project sponsor and advisors to prioritize our improvements to the interface and assessment.

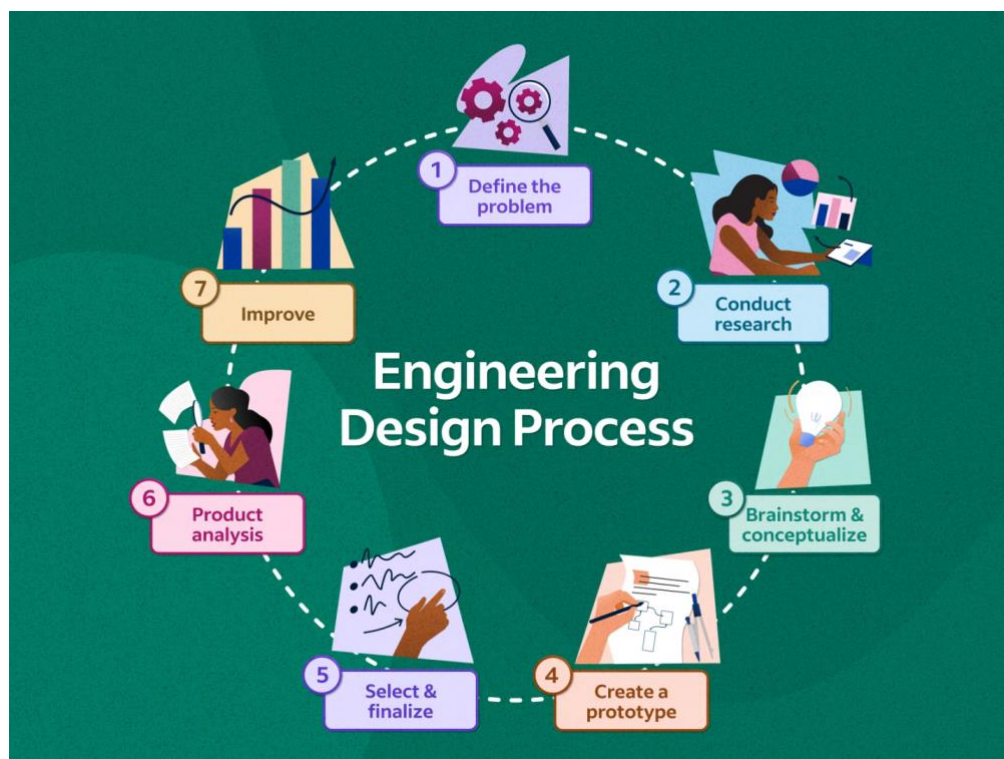


Figure 1. Engineering Design Process (*The 7 Steps of the Engineering Design Process*, 2023)

Chapter 2 will discuss the background research necessary to frame the problem at hand. Chapter 3 will formally frame the problem and discuss the design considerations taken and decisions made to address the problem. Chapter 4 will detail the methodology adopted from the considerations made in Chapter 3 to address the problem defined therein. Chapter 5 will present the results produced by the methodology of Chapter 4 both in the actual prototype and how that

prototype performed in various testing conducted. Chapter 6 will provide the team's reflection specifically on the delivered prototype, the results it produced, and potential future work, while Chapter 7 will provide the team's reflection on the project process in general.

## 2. Background

The chapter begins by introducing the social start-up, Intrare, and their work on job matching for vulnerable groups. The following sections consider the treatment of marginalized groups and the associated misrepresentation in data, the practice of skills matching, review of distinguished skill assessments, and approaches to fairness.

### 2.1 Intrare: Job Matching for Vulnerable Groups

Intrare, a social start-up based in Mexico, was founded in 2018 with the aim to promote the integration of vulnerable groups into society starting with accessible and equitable job hiring processes. The company offers an online sign-up service producing job matches for individuals of vulnerable groups based on their experience and personal information. Presently, Intrare works with refugees, migrants, single mothers, the greater Lesbian, Gay, Bisexual, Transgender, and Queer community (LGBTQI+), individuals forced to return to Mexico, and economically vulnerable women with the potential to expand their client base in the future. Intrare also partners with companies looking to diversify their talent and foster inclusivity who can post openings to the Intrare application.

Intrare is seeking a way to create more accurate matches using a skills matching approach. The resulting scores would expand upon an existing algorithm developed by Marcela Vasconcellos, a Data Science PhD candidate at Worcester Polytechnic Institute. The algorithm conducts many-to-many matching through integer optimization to maximize the fit scores for potential matches (Vasconcellos, 2023). The implementation of a pre-employment assessment could elicit important skills to factor in the larger algorithm and assist in providing more flexible roles across job titles and industries.



## **2.2 Context**

Societies worldwide have advertently and inadvertently marginalized minority demographic groups and their intersections. One way to combat this is through equitable and non-exploitative employment opportunities. While these opportunities offer great potential, matching members of these groups and their intersections to such opportunities presents its own challenges in regards to the fairness and accuracy of both the data and methods used.

### **2.2.1 Vulnerable Groups in Society**

Marginalization has to do with the relegation of minority groups to unfavorable, inadequate roles and experiences within society. The issue of marginalization is limited neither in breadth, nor in geography, nor in time; it has affected and continues to affect many demographics and their intersections in communities globally at various institutional and societal levels (Kagan & Barton, 2005). Efforts to eliminate the barriers faced by minority groups have been met with cultural and institutional backlash (Taylor, 1998). In short, barriers to free and fair societies for all are systemic, historical, global, and both actively and passively maintained wherever they exist. Overcoming these barriers requires dedicated, purposeful, and persistent action.

One such barrier is the lack of access to sustainable, equitable, and non-exploitative employment. Gainful employment is an essential aspect of societal participation, making the fair and stable employment of minority groups an essential component to their recognition as equal participants in society. Unfortunately, due to systemic discrimination, minority demographic groups and their intersections are generally disadvantaged in hiring consideration, denied offers for employment, and discriminated against while employed disproportionately compared to other

groups. Minority groups also face higher risks of exploitative and hazardous working environments as a result of their more vulnerable status in society (Green et al., 2021).

### **2.2.2 Job Matching, Data, and Analytics with Vulnerable Groups**

Job matching with vulnerable populations is paramount to the progression towards more just and equitable societies due to the role employment plays in societal integration. However, systemic biases can contaminate the data necessary for the analytical processes by which conventional job matching takes place. Indeed, these biases contaminate the data used by all analytical processes, and thus the literature on how to counteract such biases is both extensive and incredibly relevant when considering the use of analytics with disadvantaged groups (Giesecking, 2018). In short, because of systemic biases against marginalized groups, the data that is collected on them often does not accurately represent or reflect their lived experiences.

One example of this has to do with the mischaracterization of identity in data as it relates to the LGBTQI+ community. Typical data sets and collection methods define gender and sexuality as static and discrete demographic measures. Unfortunately, this approach does not accurately account for the diversity of the community nor for the complexity of individual identity as it relates to being a member of the LGBTQI+ community. To address this, some alternative identification methods have been presented. However, some of these proposed alternatives still yield problematic misrepresentations of LGBTQI+ identity. This implies that considerations of diverse populations do not necessarily address mismatches between data and reality if the considerations are not carefully constructed and implemented (Ruberg & Ruelos, 2020). This issue of mischaracterization compounds with other issues in data collection with marginalized groups, including but not limited to a lack of resources, technical expertise, and data (Ahani et al., 2021).

This problem is worsened by the importance of accurate, high-quality data in operations research and analytics applications. These applications rely on accurate data to model complex issues and problems. However, if this data is inaccurate, flawed, or misrepresentative, the results of these applications will perpetuate these issues. Especially in marginalized groups, these issues and biases contained in data can exacerbate their issues with marginalization, causing greater harm to vulnerable groups. The aforementioned data collection issues in conjunction with the need for responsiveness in humanitarian operations make the data collection and utilization, and analytics challenging to implement in a humanitarian context

That said, applications of analytics to such contexts present incredible opportunities in addressing this issue. In their study of the role of data in achieving equity, Chauhan and Kshetri note that analytics can be used to identify hidden trends related to data disaggregation and diversity efforts (Chauhan & Kshetri, 2022). They also note specific successes in the use and application of data, including identifying issues faced by certain marginalized communities such as women's employment during the COVID-19 pandemic and the treatment of sickle cell disease in populations of African descent (Chauhan & Kshetri, 2022). However, even this positive outlook on the use of data and analytics with marginalized populations notes the risks and considerations mentioned above (Chauhan & Kshetri, 2022). As such, analytical processes can work to remedy the issues of marginalization, but only when carefully and intentionally implemented (*Analysis and Recommendations*, 2021).

### **2.2.3 Recommended Mitigation Techniques**

To combat the potential issues of data misrepresentation and misclassification of vulnerable groups, some mitigation techniques are recommended by field experts. Specifically noted risks and challenges related to data collection include apprehension about the disclosing of

certain information, gaps in self-evaluated responses and the reality of certain situations, and the relationship between data collectors and the sample population. To overcome these risks and challenges, experts recommend making assurances that underrepresented groups are represented in data, that data collection makes a concerted effort to avoid harmful actions, and to identify and eliminate potential exclusions (*Analysis and Recommendations*, 2021).

## **2.3 Skills Matching**

The modern job market has shifted to a more skills-focused view on jobs and candidates. Companies are using skills assessments to accurately measure the skills of potential employees. Using these assessments, they can evaluate their soft and hard skills to create better and fair job matches.

### **2.3.1 Skills Matching Overview**

As the job market is increasingly specialized in the modern era, candidate skills are essential. A survey of over 1,000 workers and 200 business and Human Resources (HR) executives found that 89% of executives believe skills are important in defining talent, deploying talent, managing careers, and valuing employees (Jooss et al., 2023). With the current shift from staffing for jobs to staffing for skills, HR is then able to use skills assessments to evaluate potential job candidates more objectively and find the best fitting employee for a job (Jooss et al., 2023). Skills matching is “the process of comparing job opportunities with prospective applicants, typically culminating in a ranked list of recommendations” (Bogen & Rieke, 2018). Companies are using assessments like questions, video interview analysis, and gameplay to aid in getting accurate and in-depth results (Raghavan et al., 2020). Companies can use these platforms to assess the skills of

potential employees. The majority of the assessments categorize the skills into hard and soft skills. With a proper list of the jobs' required skills, HR can then match job candidates to open jobs.

### **2.3.2 Hard versus Soft Skills**

Hard skills are the technical abilities related to specific tasks or functions for a job. Acquiring skills in programming, graphic design, and carpentry typically involves training and education. Hard skills are important for specific tasks in an industry that requires expertise (Lamri & Lubart, 2023). They are easily assessed using tests specific to the skill like written and practical tests.

Soft skills are defined as behavioral skills such as communication, teamwork, and time management. To gain or improve on these skills often requires dedication, self-reflection, and self-improvement (Lamri & Lubart, 2023). During the development of one soft skills assessment, the most important soft skills in the work setting were problem solving, teamwork, communication, time management, and decision making (Altomari et al., 2023). Soft skills, unlike hard skills, are difficult to measure and quantify. However, there are assessments dedicated to specific soft skills. For example, to assess teamwork, the Comprehensive Assessment of Team Member Effectiveness (CATME) evaluates users' various levels of performance in CATME's five teamwork dimensions. The five teamwork dimensions are "contributing to the team's work, interacting with teammates, keeping the team on track, expecting quality, and having relevant KSAs (Knowledge, Skills, and Abilities)" (*Welcome to CATME - Smarter Teamwork*, 2023). These skills are assessed by teammates on a scale from one to five and results of the assessment include scores of self, team, and average. Additionally, using the General Decision-Making Styles (GDMS) scale people can test their decision making. It distinguishes between five decision styles: rational, avoidant, dependent, intuitive, and spontaneous. The assessment has 25 questions using a five-point rating

from one being strongly disagree to five being strongly agree (Appelt et al., 2011). Assessments like these are frequently being designed and developed to help in hiring, training, and education.

### **2.3.3 Analysis of Benefits and Drawbacks**

Skills matching provides benefits to the company and employees because it reduces turnover and helps with career development (Jooss et al., 2023). An effective skills matching program allows for more accurate pairings of candidates to companies, which helps increase employee longevity and reduce turnover. With accurate results, individuals can understand potential skill gaps and improve on weaker skills (Hane-Weijman, 2021). For example, OpenSKIMR, a project funded by the European Union, is a platform used to match users to career opportunities and suggests future training. The program assesses the user's skills using an algorithm to create job matches. A second algorithm evaluates the gap between a candidate's skills and those required for the desired job. With this knowledge, it can suggest a plan for improving their skills gap (Rentzsch & Staneva, 2020). This holistic approach to skills assessment and feedback allows for improved matches and career development.

Self assessments allow for users to inaccurately score themselves leaving a gap in results. A study conducted at three companies compared employees' and employers' assessment of their soft skills, and found that there were perceived differences in the employees' soft skills between the employee and employer (Tsirkas et al., 2020). The potential accuracy gap with self-reported soft skills assessments should be considered when reviewing these results.

## **2.4 A Review of Prominent Skills Assessments**

Pymetrics and Koru are prominent vendors of skills assessments; however, their practices introduce the opportunity for bias during the hiring process (Bogen & Rieke, 2018). The proposed

platform NADINE attempts to address the needs of protected groups with their skills assessment.

### 2.4.1 Pymetrics

Pymetrics is a prominent vendor of soft-skills assessments that has risen in popularity in recent years, being used by companies such as JPMorgan, Kraft Heinz, and Swarovski. Their assessments use neurological games backed by behavioral science to assess candidates' effort, risk tolerance, decision making, attention, focus, learning, fairness, generosity, and emotion (“Soft Skills Assessment Testing – Pymetrics”, 2023). One such game is the balloon analogue risk test (Figure 2), where participants are asked to pump up a balloon, earning a reward that increases in value for every pump. However, if the balloon is inflated past a variable threshold, the balloon pops, and the reward is lost. While the objective of the game appears to be to earn the most rewards, it is actually testing the risk tolerance of the candidate (Lejuez CW, 2002). For each company that uses Pymetrics, a specialized predictive model is built using the game performance data of successful employees at that company. Prospective employees are compared against the successful employees to assess their fit in the role (“Pymetrics Internal Demo Day Pitch”, 2017).

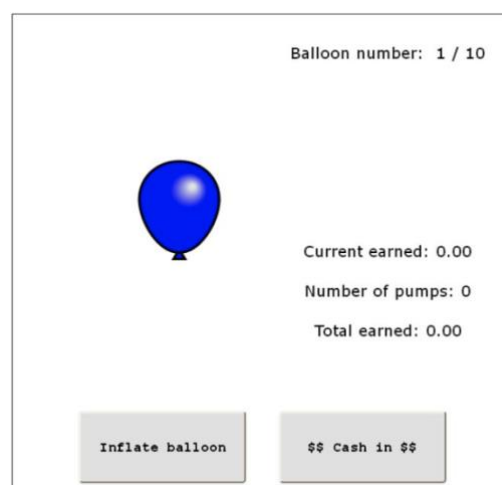


Figure 2. Balloon Analogue Risk Test (Lejuez et al., 2002)

Pymetrics' approach of comparing applicants to successful employees presents an opportunity for discrimination. As discussed previously, employee evaluations are known for being inaccurate, subjective, and biased (Bogen & Rieke, 2018). Subjective evaluations are a common source of discrimination, and models derived from these evaluations can mirror their discriminatory patterns (Bogen & Rieke, 2018). A possible source of bias in the evaluations is the identification of positive traits. The positive traits in the evaluation model may not be correlated to success, but rather correlated to manager satisfaction, leading to inaccurate predictions of candidates' abilities. Additionally, by looking for specific traits, candidates who do not meet the profile of existing employees may be turned away, despite being qualified for the position.

Biases in the field of psychology may also influence the accuracy of Pymetrics' games. For example, many psychological studies rely on college students as their primary research subjects, so their results do not always apply to less educated individuals (Henrich et al., 2010).

### **2.4.2 Koru**

Koru is an alternative to Pymetrics that assesses potential employees based on personality traits. Their product poses either-or questions to candidates, who then select the option that most aligns with their sense of self. The results of the quiz are used to assess the candidate's fit for the prospective employer. Similar to Pymetrics, Koru determines fit using the assessment results from existing employees (*Predictive Hiring*, 2018). For each potential candidate, Koru provides a fit score as well as individual scores for the assessed traits (*Predictive Hiring*, 2018).

Koru has many of the same potential biases as Pymetrics, but the fit score that Koru provides to employers presents an opportunity for biased decision making. If the score is used to rank candidates, recruiters may interpret nominal score differences as implying significant differences between applicants. However, the competencies of two similarly scored candidates are



likely close to equivalent. Employers may also give the rankings undue significance, using it as a highly accurate competency marker instead of a screening score. Additionally, if certain demographic groups perform slightly better on the assessment, they may be selected for the role more frequently. This can lead to an increased disparity between groups, since members of that demographic group will not be included in the target data set.

### **2.4.3 NADINE**

The proposed NADINE soft skills assessment aims to address the needs of migrants, refugees, and asylum seekers in Europe. It is a questionnaire that assesses thirty-five skills such as cognitive ability, numeracy, and communication. Following the completion of the assessment, a skill vector is available to the user and can be added to their resume. The NADINE skills vector is not publicly available, but a skills vector from a different vendor is shown below to illustrate the concept (Figure 3). In contrast to Koru and Pymetrics, NADINE uses the skills assessment to recommend job categories that align with the user's abilities. NADINE also adds additional functionality by allowing host authorities or NGOs to use the skills vector to recommend cities that would be ideal for a displaced person to relocate to (Leligou et al., 2021).

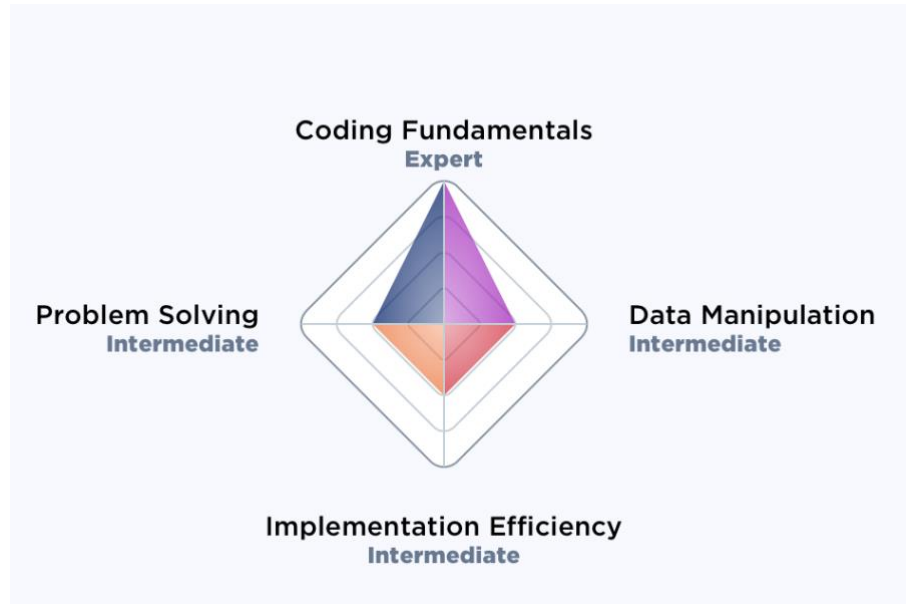


Figure 3. Alternative Skills Vector (CodeSignal, 2023)

## 2.5 Implementation of Fairness in Algorithms

When working with disadvantaged groups, it is critical to consider the long-term impacts and the potential for unfair outcomes. The job hiring domain has a history of discriminating against vulnerable groups such as women and minorities (Chirica, 2021). Algorithmic use within the field has led to discussion on the implications of fairness and the potential to perpetuate or create new biases (Raghavan et al., 2020). A thorough understanding of current fairness practices and the potential to cause harm in a sensitive field is necessary to implement beneficial solutions.

### 2.5.1 Overview of Fairness

Fairness is a highly debated topic within the algorithmic world, and there is no universal definition of fairness. Instead, the definition differs depending on the context of the application (Lewicki et al., 2023). While this provides more flexibility, it limits the ability of researchers to determine what is or is not fair. There is a recognized trade-off between algorithmic fairness and utility where the balance is also highly contextual. Algorithms that do not implement fairness

constraints can achieve higher levels of performance. However, in many instances, fairness is favored over optimality to prevent biased outcomes and increase utility (Bertsimas et al., 2011). While choosing a model based on predictive performance can be more biased than alternatives, an intense focus on fairness can skew the results such that they are no longer accurate and could be considered unfair (Lewicki et al., 2023). The balance between these two extremes is possible but requires careful implementation to achieve the target outcome.

Fairness can be achieved in various ways, at different stages in the algorithmic process to reduce bias in models. One approach, fairness through unawareness, adjusts the training data by removing sensitive features found to have harmful impacts on the model outcome (Raghavan et al., 2020). Yet removing this sensitive information makes monitoring equal outcomes difficult and creates the potential for more bias, especially for more vulnerable groups. In one study, an employee's distance from work related to their tenure; however, this distance exhibited a prominent correlation to race (Bogen & Rieke, 2018). This study exemplifies how results can be misconstrued and unknowingly perpetuate socioeconomic disparities.

Fairness can also be enforced as an algorithmic constraint. For example, individual fairness aims to produce similar outcomes for similar people, while group fairness looks at the overall large-scale fairness of demographic groups in various implementations. Some of these approaches include obtaining similar outputs or balancing error rates across different groups. However, individual and group fairness focuses more on equating equal outputs than the resulting real-world impacts (Finocchiaro et al., 2021). Another approach, within-group fairness, compares the model's results for differing groups and selects a top percentage from each category. This evaluates the outcome's fairness to ensure more diversity within selection processes (Raghavan et al., 2020).

## **2.5.2 Applications within Algorithms**

Many machine learning algorithms rely on training data to predict outcomes. This approach allows the model to learn from itself, increasing its accuracy through prolonged use. The foundation of machine learning processes relies on identifying relationships to predict outputs from input data. This concept fails to acknowledge the impacts of misclassifications in the system and the potential to amplify harm as the model continues to train itself (Finocchiaro et al., 2021). If the potential for biases are not considered, it is unclear how fair a model can be.

Another consideration for algorithmic implementation is the potential to game the system. Users can optimize their inputs for desired results, which causes inaccurate results that skew the training model. Multiple approaches exist to limit the impacts of strategic behavior on model validity. One example is utilizing incentive-aware machine learning in algorithms by implementing strategic classification to prevent these misclassifications (Chen et al., 2019). Forced choice is another technique utilized to prevent strategic behavior. When users are forced to rank their answers, they can more fairly consider and present both their strengths and weaknesses. This reduces the need to choose the supposed optimal answer and results in more accurate results (Bartram, 2007).

## **2.5.3 Algorithmic Fairness in Action: Skills Matching**

There are numerous applications of algorithmic use within the job hiring domain including the implementation of skills intelligence and matching. One practice, labor market analysis, tracks the supply and demand of various skills to keep up with the ever-changing needs of the labor market. To detect and predict these trends, labor market analysis often pulls from national databases such as the Occupational Information Network (O\*NET) and the European Skills, Competences, Qualifications and Occupations (ESCO). One company that implements these

practices is SkillLab, a Dutch start-up assisting refugees seeking employment. SkillLab generates customized materials based on the user's experience, background, and relevant skills with the intent to create stronger candidates in the hiring process. The ultimate goal is to ensure social and economic equality among refugees through sustainable and applicable employment (Rentzsch & Staneva, 2020).

#### **2.5.4 Further considerations**

The implications and accessibility of fair job hiring have critical applications on individuals and society. Human-based processes are more procedurally fair, but applicants view the use of algorithms in hiring as less so. Users' negative perception of algorithms and the impact on the hiring processes reflects back on the employers. This can impact companies' reputations thus reducing the number of applicants and acceptance offers (Lavanchy et al., 2023).

Despite an algorithm's ability to maintain fairness and accuracy, the way an employer or recruiter will interpret that information is unclear. For example, there is an unknown impact on how rank-ordered lists and numerical scores can impact recruiters' choices. This means the results of these algorithms can be misinterpreted and result in biased outcomes despite the initial considerations on fairness (Bogen & Rieke, 2018).

The complexity and highly dependent ties to the context of the problem makes algorithmic fairness difficult to understand and implement. However, with the potential to create positive change, it is imperative to recognize the potential to create bias and additional harm.

## **3. Design**

The goal of our project is to enhance Intrare's job matching algorithm with a soft skills assessment and corresponding interface. We first defined the problem and considered constraints. We then designed a soft skills assessment based on information from Intrare, a review of successful questionnaires, and consideration of best practices. The interface of the assessment was designed with the principles of universal design and the Intrare brand manual (Persson et al., 2014; Intrare, personal communication, 2023).

### **3.1 Defining the Problem Statement**

This section will further consider the problem and what steps were taken to produce a skills assessment. We outline necessary requirements including ensuring accessibility, maintaining accuracy, and limiting biases to align with specifications of the project. Next, we considered the constraints of the project and associated risk mitigation. Then, we investigated the definition of success in the project to ensure we can accurately and effectively complete the project goal.

#### **3.1.1 Problem: Intrare**

This project aimed to answer three main questions faced by the proposed assessment. The first aspect we had to understand is what traits, skills and abilities make sense for a match. We also needed to consider how to elicit that information from users. Finally, the ability to integrate the results of the assessment into the existing system was crucial for accurate matches. By addressing these three questions, we sought to create an assessment that successfully elicits information to be integrated into the established platform.

### **3.1.2 Project Requirements**

These specifications were developed based on conversations with Intrare about their organizational values, the existing application, and other considerations for the creation of a skills-matching assessment including the opinions of Intrare users. The main overarching requirements were to ensure accessibility, maintain accuracy, and limit biases.

Accessibility was necessary to provide the service to a vast, diverse number of individuals and to avoid implementing additional barriers. Accessibility was intertwined with every stage of the project from the skills selection to the questionnaire to the interactions with the interface. For example, the questions were crafted to be easily understood and answered while maintaining audience engagement. One consideration was the clarity of the proposed questions, and the potential to lose comprehension when the questions were translated from English into Spanish.

Another project requirement was maintaining a high level of accuracy while limiting biases. When working with vulnerable groups, there is potential to perpetuate or spread new biases (Lewicki et al., 2023). However, we also wanted to ensure that the skills-matching would strengthen the existing system, and the skills were comprehensive of the target populations. There needed to be consideration of where to integrate fairness and why.

### **3.1.3 Constraints**

In meeting these requirements, it was also imperative to consider the anticipated limitations of the project's scope. The project time frame of six months especially limited the possible ways in which the requirements were met. Risk mitigation also had to be built into the project. The greatest risks to this project come in the form of feedback and lack thereof regarding the final deliverable. We anticipated that obtaining feedback would enhance the time constraint, as collecting and iterating on feedback would take time that could otherwise be devoted to the design

and implementation of our deliverable. The combination of these constraints pushed the group to account for potential buffer time in our production schedule. This consideration was mainly to allow for getting feedback and improving the deliverable. The group also expected to prioritize the design considerations according to their utility to achieve the project goals. We rendered several design considerations infeasible within the scope of this project. While some designs were satisfactory, they would exclude considerations more essential to the project.

### **3.1.4 Project Goals**

The objective of this project was to augment the existing job matching process at Intrare by incorporating candidate skills data. To achieve this, we developed a skills assessment questionnaire and an interface through which the questionnaire is administered. The questionnaire and its interface were delivered to Intrare before the project's end date on March 1st, 2024. Successfully completing this objective better informs the augmented matching algorithm, allowing it to produce more accurate matches between Intrare talent and partner job offers. These more accurate matches in turn give Intrare users a higher likelihood of being matched to job opportunities where they have better chances of succeeding and more sustainably maintaining their employment. This success will also increase the utility of Intrare to its partner employers, as these employers would be able to more easily find job candidates with the potential for success in the employer's open positions. These results in tandem will make Intrare a more appealing platform both for diverse talent and for employers.

## **3.2 Designing the Assessment**

We aimed to identify key parts of other skills-based assessments and combine them to create our own assessment. The first stage of the process was designing the assessment's parts;



which were soft skills, assessment type, scoring, and accessibility. We developed the soft skills from Intrare’s database, similar assessments, and previous research. Options that we considered for the assessment type were questionnaires, games, and VR. When looking at each assessment, we considered the limitations each poses and the biases that they can possess. We then reviewed the factors we plan to use in creating questions for the assessment. Using the selected type of questionnaire we finalized the scoring we will use for the assessment. Finally, we considered how we could create an accessible assessment for users.

### 3.2.1 Creating a Comprehensive Skill Set

To develop a soft skills assessment, we first needed to create a list of skills to guide our assessment. During this process, we developed a list of skills from Intrare’s database, similar platforms, and research. While conducting research, we found most assessments test soft skills from a wide range of categories to cover all traits of the user. Specifically, the NADINE project test users over four categories, shown in Table 2. With guidance from Intrare to keep the assessment between 10 and 15 minutes, we decided to assess 5 to 7 soft skills. Using the list of necessary skills from Intrare’s job database, we created a frequency chart to evaluate the most common skills, shown below in Table 1.

*Table 1.* Intrare Skills Frequency Chart

<b>Soft Skill</b>	<b>Frequency</b>
Dynamic	16.13%
Attention to detail	16.13%
Communication	12.90%
Teamwork	9.68%
Empathy	9.68%
Proactivity	6.45%

Leadership	3.23%
Integrity	3.23%
Decision Making	3.23%
Cooperation	3.23%
Responsibility	3.23%
Learning Capacity	3.23%
Commitment	3.23%
Learning Ability	3.23%
Presentation Skills	3.23%

Then, with the use of other skills assessments with similar missions to Intrare, we were able to further develop the list of soft skills. We looked at the proposed NADINE project, which assesses 19 soft skills over four categories using a questionnaire. Below in Table 2 are the four categories with their respective soft skills (Tountopoulou et al., 2020).

*Table 2. NADINE Soft Skills*

Interpersonal skills:	Social interaction, Teamwork, Intercultural competence, Coping with Authority, Conscientiousness, Extraversion
Organizational Skills:	Work efficiency, Time management, Independent Work, Organization
Personal Skills:	Adaptability, Reliability, Willingness to learn, Stress tolerance
Entrepreneurial skills:	Creativity, Managerial skills, Initiative, Risk tolerance, Leadership

Finally, we compiled a list of important soft skills in the workplace based on research. We used a 2011 survey that asked 49 executives to list the ten most important soft skills. The results of the study were as follows in Table 3 (Robles, 2012).

Table 3. Top 10 Workplace Soft Skills

10 Soft Skills	<ul style="list-style-type: none"><li>● Integrity</li><li>● Communication</li><li>● Courtesy</li><li>● Responsibility</li><li>● Social skills</li><li>● Positive attitude</li><li>● Professionalism</li><li>● Flexibility</li><li>● Teamwork</li><li>● Work ethics</li></ul>
----------------	--

Considering all three sets of soft skills, we combined and developed our set of soft skills: teamwork, communication, organization, adaptability, leadership, and willingness to learn. This set was then reviewed with the entire team including Intrare which was adapted to: teamwork, communication, organization, adaptability, intellection, and diligence/ownership. Through discussion with Intrare, intellection replaced willingness to learn. Secondly, diligence was selected over leadership because leadership may be hard to measure and diligence aligns more with responsibility. With the final list of soft skills, we were then able to start creating and developing questions to elicit the skills.

### 3.2.2 Limiting Bias in Pre-Employment Assessment Type

Pre-employment assessments are utilized early in the hiring process and closely impact the later hiring pipeline. Pre-employment assessments have historically unfavorably impacted minorities, requiring a closer look at the validity and impacts of varying approaches. Gameplay and video interviews recently grew in popularity as the more modern and innovative alternative to questionnaires. Yet, the techniques to validate the results of these assessments are rarely available, calling into question the accuracy of results and the potential for biases. Video game assessments

specifically rely on facial analysis yet the error rates vary by gender and race. It is also unknown if facial expressions can infer emotions especially with respect to cultural differences and individuals with disabilities (Raghavan et al., 2020).

These newer approaches may be more intriguing, yet the potential for further bias inhibits the use in the context of this project. Instead, we aimed to combine some of the interactive elements present in games with the scientifically-backed questionnaire. As mentioned earlier, there are varied approaches to questionnaires providing more flexibility and similar applications from which to learn (Raghavan et al., 2020). While developing a game had the potential to be more interactive, conversations with our advisors and Intrare solidified that a questionnaire aligns best with the project needs and the focus on prohibiting biases within the available timespan.

### **3.2.3 Develop Questions for Assessment**

To successfully develop a questionnaire, we needed to further break down the design and development of the assessment. Our group used the model for survey development included below to guide the process and ensure no aspects are overlooked (Strachota et al., 2013). With the literature review and problem statement developed, we moved to the later stages centered around the construction and evolution of the questionnaire including the questionnaire type, questions list, and the final number of questions. By analyzing each of these aspects, we assembled a questionnaire with large consideration to the context of our project and the target groups.

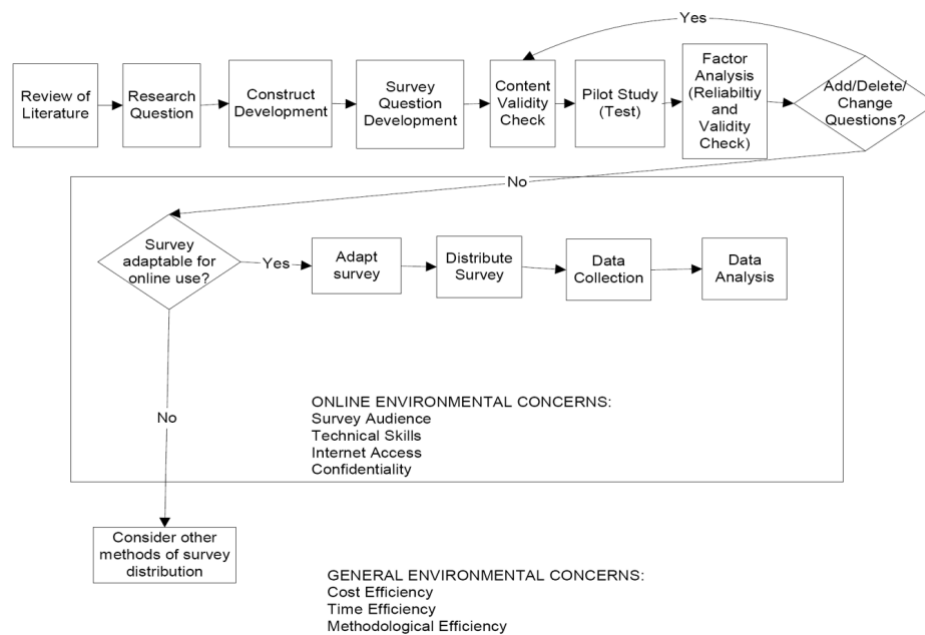


Figure 4. Model for Survey Development (Strachota et al., 2013)

Based on previous research and similar soft skills questionnaires, we decided to investigate open and closed questions. Open questions allow for more in depth responses from the user. The downside of this type of question is the ability to analyze and interpret the responses. While closed questions are more restrictive, they are easier to interpret and compare scores from a large number of respondents (Rattray & Jones, 2007). With this in mind, we decided to develop our assessment using closed questions.

A common closed question type in questionnaires is the Likert scale. The Likert scale is a psychometric scale with categories for users to respond with their opinions, attitudes, and feelings toward issues. The scales range from zero (strongly disagree) to five (strongly agree). The Likert scale's advantage over other questionnaires is that it allows for a large amount of accurate data to be collected and easily compared (Nemoto & Beglar, 2014). Varying individual interpretations of the scale and tendencies to avoid extreme answers can impact the results of Likert scale questions

(Heo et al., 2022). Yet, this type of questionnaire structure allows for an intuitive self assessment (Al-Sa'di et al., 2023).

After the Likert scale was selected as a basis for the questionnaire, the next stage of our design was to develop the questions for the assessment. There was little available information on developing questionnaires within the context of our project. However, there are a multitude of surveys targeted at understanding the soft skills of educators and students (Al-Sa'di et al., 2023; Nemoto & Beglar, 2014). One questionnaire aimed to compare the soft skills of educators before and after the COVID-19 pandemic. In this assessment, during the development stage, they started by conducting a literature review and selecting the soft skills, identifying independent variables, and then worked through assembling the questionnaire. The questions for this assessment were crafted during multiple focus groups and utilized the Likert scale to collect responses (Al-Sa'di et al., 2023). This similar application solidifies the choice to use a Likert scale questionnaire and provides guidance on how to develop our questionnaire.

Because Likert scale assessments are construct-centered, they are typically built around the context of a particular hypothesis or question. In relation, each question should be centered around measuring one element of the concept (Nemoto & Beglar, 2014). The process for developing the questions in a Likert scale context aligns with the best practices for traditional survey development (Nemoto & Beglar, 2014; Strachota et al., 2013). Some of these practices include providing straightforward and easy to understand questions, avoiding the use of conjunctions (Nemoto & Beglar, 2014), vagueness, multipurpose or questions with multiple answers, manipulative or biased wording, use of typography (Strachota et al., 2013), and avoiding conflicting questions (Rattray & Jones, 2007).

Another element of developing the survey was determining the number of questions to include on the assessment. For Likert scale questionnaires, it is recommended to pilot 10 to 12 questions to ensure there are 6 to 8 questions to accurately capture each concept. (Nemoto & Belgar, 2014). In this project, with six approved skills and the above recommendations, our group planned to have 60-72 questions to pilot and ideally would cut down to 36-48 total questions. However, a main consideration of the project is limiting the survey's time requirement to increase accessibility. During conversations with Intrare, the goal was to keep the total time of the questionnaire between 10 and 15 minutes. The need to include enough questions to produce accurate results and meet the outlined time suggestions became a key consideration in developing the assessment.

After the creation of the questions for the soft skills assessment, the questions were then developed using subject matter expertise. This process of editing and refining questions for skills assessments is very common. During the creation of a soft skills questionnaire for nurses, the developers used subject matter experts to evaluate questions. Using their own rating system the experts were able to remove and edit the starting list of 30 questions to 25 final questions (Aridi et al., 2023). The development of a similar assessment, SKILLS-in-ONE, also closely involved experts to create and develop questions. They utilized psychologists and expert evaluators to review their questionnaire items. With this process, they were able to find 13 questions that needed to be reworked and then put back into the item list (Escolà-Gascón & Gallifa, 2022). With the use of experts, both soft skills assessments were able evaluate and develop the questions for their assessment.

In the context of the proposed project with Intrare, the experts to evaluate questions were Intrare employees. During this process, the experts rated and made suggestions for our proposed

questions. With this feedback, the soft skills questions were developed to elicit the most accurate information from employees.

### **3.2.4 Scoring of Assessment**

One approach to assigning a skills match score is detailed by Petrican et al., and uses an ontology of skills based on descending hierarchy to produce a single match score based on the lowest common ancestor of each candidate skill and each job offer skill (Petrican et al, 2017). While this approach has several documented advantages, the context of our problem makes such an approach infeasible. Firstly, the ontology proposed would be difficult to produce within our context of primarily soft skills, whose hierarchical relationships are subjective and poorly defined. The approach is also best suited for a vast, varied database of skills assessed, which is not suitable in this context due to the relative similarity and focused nature of the skills demanded by job offers. Lastly, this approach assumes a binary presence of a given set of skills and uses this presence to yield a single similarity score between a candidate and a job offer. This approach fails to accomplish two objectives of our own solution, namely that of providing a scaled assessment of the presence of a given skill in a certain candidate and that of an individual score for each candidate in each skill assessed to provide the candidate insight into their own capabilities.

Another alternative approach develops a knowledge graph of the Jaccard distance between skills and occupations based on their relationships according to the International Standard Classification of Occupations (ISCO) and ESCO taxonomies (de Groot et al., 2021). This approach is more applicable than the previous with soft skills because the relationships between skills and occupations are better defined within the developed dataset. Unfortunately, the approach maintains the issues of yielding a single match score dependent on job offers as opposed to presentable candidate skill scores and of utilizing a broad, diverse set of skills. Additionally, this approach has



a unique issue making it infeasible for the context of this specific problem. Namely, it relies heavily on candidate résumé data in order to generate a skills match, making it infeasible in the context of Intrare, where candidate data is often incomplete or unavailable. This would render the approach impossible for candidates without résumé data, which in the case of Intrare would be a plurality of use cases.

The NADINE skills assessment discussed in Section 2.3.3 utilizes a similar Likert scale to the one that we will be using. The NADINE raw score is generated by adding all of the answers together for each skill (Rodriguez et al., 2020). This approach is standard when working with Likert scales (Likert, 1932). There are proposed changes to the traditional summation system, such as the weighting proposed by (Chakrabarty, 2014). However, many of these techniques rely on data from candidates who have already taken the assessment. Because this is a pilot, we will not have previous assessment data, and cannot implement these techniques.

### **3.2.5 Designing an Accessible Platform for the Assessment**

We created a wireframe before implementing our questionnaire to ensure the accessibility of our user interface. This wireframe needed to use principles of accessible web design as well as general principles of aesthetic web design. After we created the wireframe, we used it as a template for implementing our final questionnaire on the Internet as part of Intrare's platform.

First, we considered the principles of accessible web design. There is no single definition for accessibility as it relates to web design, making its implementation open-ended and dependent on developer interpretation. For our wireframe, we used the seven principles of universal design as developed by Ronald Mace (Persson et al., 2014). Specifically, these principles are equitable use; flexibility of use; simple, intuitive use; perceptible information; tolerance for user error; low physical effort; and size and space for approach and use (Persson et al., 2014). We focused

specifically on the third of these principles: simple and intuitive use because of the target user base. This principle calls for the product being easy to use regardless of the intelligence, experience, language skills, and concentration level of the person using it.

Our wireframe had to adhere to the principles of good web design. We aimed to make it easy to read and navigate through, with the important information put clearly. We also carefully considered the wireframe's appearance when viewed on mobile devices. The wireframe also had to comply with the Intrare brand manual outlining the company's design practices (Intrare, personal communication, 2023). A draft of this wireframe is shown below in Figures 5, 6, and 7:

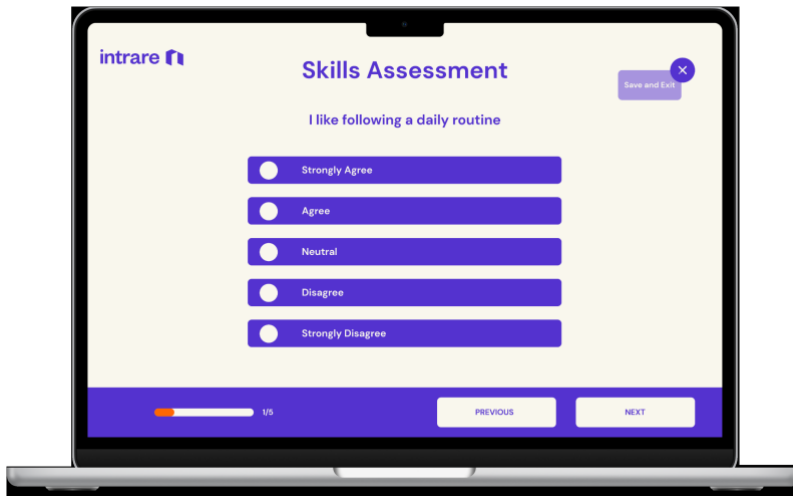


Figure 5. First Draft of Wireframe Assessment

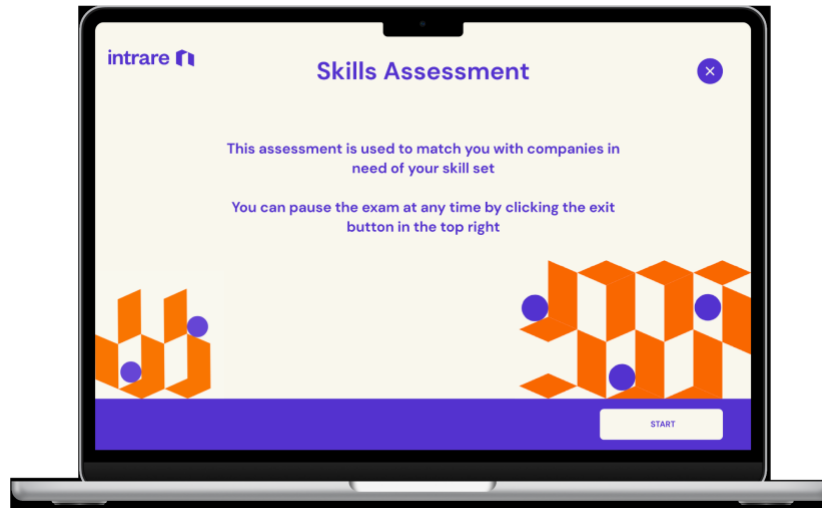


Figure 6. First Draft of Wireframe Home Page



Figure 7. First Draft of Wireframe on Mobile Phone

## 3.3 Evaluating the Assessment and Interface

After the assessment and interface were developed, we assessed their success through evaluations with Intrare users and employees. Because Intrare employees are more experienced with providing feedback on interfaces, we used the usability heuristics model introduced by Nieslen and Molich (1990). With Intrare users, we planned to gather feedback with a questionnaire based on guidelines published by the Nielsen Norman Group (2023) because it demands less of the user's time than a focus group or cognitive walkthrough.

### 3.3.1 Interface Testing

We conducted user testing with the Intrare team members to evaluate the interface. There are a number of well established user testing methods such as heuristic evaluations, cognitive walkthroughs, and focus groups (Rohrer, 2022). We evaluated the interface with the Intrare team using the usability heuristics model introduced by Nieslen and Molich (1990) because this technique is ideal when working with subject matter experts, while the other methods are ideal for end users. We had three to five Intrare employees evaluate the interface, serving as the subject matter experts in this study. We also used the severity ratings defined by Nielsen and Mack (1994).

**0** = I don't agree that this is a usability problem at all

**1** = Cosmetic problem only: need not be fixed unless extra time is available on project

**2** = Minor usability problem: fixing this should be given low priority

**3** = Major usability problem: important to fix, so should be given high priority

**4** = Usability catastrophe: imperative to fix this before product can be released

The literature review by Hermawati and Lawson (2016), included heuristic ratings for a variety of domains. Sim (2008) provides recommended heuristics for computer assisted

assessments. We adjusted these heuristics to be applicable to our application, but they are largely the same and consist of heuristics 1-11 in Table 4. We have also added the heuristics from Díaz-Arancibia (2013) as heuristics 12 and 13.

*Table 4. Heuristics and Descriptions to Be Used in the Heuristic Evaluation*

<b>Heuristics</b>	<b>Description</b>
1. There is clear language and grammar within questions	Text should be grammatically correct and make sense.
2. Progress through the test is visible and understandable	Ensure that the number of questions answered and remaining is obvious and time remaining is clear.
3. Answering questions is intuitive	Clear distinction between question styles and the process of answering the question should not be demanding. Answering the question should be matched to interface components.
4. Actions are easily reversed	It should be possible to change or remove an answer. Ensure it is possible to return to an incomplete test or question.
5. Users are informed of any unanswered questions before finishing	If a user has opted to end the test ensure that they are informed of any unanswered questions.
6. Interface design characteristics are appropriate	Interface should match standards and design should support user tasks.
7. Visual layout is appropriate - adequate spacing and visibility of questions	Ensure that there is enough spacing between the elements within the interface and scrolling is minimized within the questions.
8. Feedback is appropriate	System feedback should be clear about what action is required. Question feedback should assist the learning process.

9. Moving between questions and terminating the exam is intuitive	User input to navigating between questions and returning to unanswered questions should be consistent. Options to exit should be identifiable.
10. Time delays are minimized	Prevent any unnecessary delays. Ensure that there is minimal latency when moving between questions or saving answers.
11. External influences to the user are minimized	Ensure test mode does not impact on fairness and performance within the test. Prevent distractions to other users and do not penalize them due to constraints of the software e.g. spelling mistakes (unless essential)
12. The system matches the real world	Independent of the objective culture, the website should speak the language of its users with words, phrases and concepts familiar to them, making the possible actions easy to understand, instead of using concepts related to the system or technical terms. The website should use conventions of the objective culture such as: images, colors, and familiar objects, in a logical and natural order.
13. Users can realize things in a consistent and standard way	Independent of the objective culture, the system should not make the users think if two actions, situations, or different words mean the same. Established conventions must be followed, under conditions that the user can realize things in a familiar, standard, and consistent way.

After the Intrare employees participating in the evaluation rank each heuristic on the severity scale, they will discuss the ratings amongst themselves, producing a list of issues with the interface ranked by severity. We will then use this list to prioritize improvements to the application.

### **3.3.2 Assessment testing**

Following the heuristic testing with Intrare employees, we planned to administer a survey to Intrare users to evaluate the user experience. We proposed to use a questionnaire as opposed to

a cognitive walkthrough or focus group since it demands less of the users time. This is an important consideration as we are working with vulnerable groups with competing commitments. This aspect of testing was planned to focus on our assessment, while we assessed the interface. We developed the questionnaire based on guidelines published by the Nielsen Norman Group (2023). The questionnaire is displayed in Table 5. While this wasn't completed within the project timeline, it is recommended as a future work.

If working with clients who have already been matched to employers, it would be prudent to evaluate them again based on the assessment. This would enable validation of the results of our assessment.

*Table 5. Questions for the User Questionnaire*

<b>Question</b>	<b>Possible Answers</b>
1. This assessment was too long	Strongly disagree, disagree, neutral, agree, strongly agree
2. This assessment was too short	Strongly disagree, disagree, neutral, agree, strongly agree
3. I understood what the questions meant	Strongly disagree, disagree, neutral, agree, strongly agree
4. I understand my result	Strongly disagree, disagree, neutral, agree, strongly agree
5. This result represents me well	Strongly disagree, disagree, neutral, agree, strongly agree

## **4. Methodology**

Building off our proposed design, this section will further outline the evolution of the assessment. We start with an overview of the development of the questions based on the selected skills. Next, the front end development is discussed to highlight the shift from the prototype to product. We will then examine the back end development before looking at the skills scoring vector to display the user's results from the assessment.

### **4.1 Questions**

The team underwent multiple versions of questions to ensure that we maintained accuracy and accessibility. First, we further researched each skill and similar assessments to create an initial draft of the questions. We then reviewed with Intrare, advisors, and experts to rephrase the questions for clarity, while maintaining their underlying intent. We also discuss how we selected the order of the questions and the choice to implement a milestone system.

#### **4.1.1 Original Draft Research**

To best evaluate the given skills of potential candidates, the team conducted an investigation into the nature of the skills to be assessed. Specifically, it was necessary to identify both the components of having strong adaptability, communication, diligence, intellection, organization, and teamwork skills, as well as the traits someone with such skills would exhibit. Using this research and questions developed therein as a basis, as well as other questions from similar publications, an initial list of questions was compiled. Based on the nature of the target user base for the questionnaire and through cooperation with Intrare, it was determined that the list of questions should avoid negative phrasings to encourage greater user participation and avoid



intimidating users. Additional attention was paid to ensure the items in the questionnaire were relevant to the users for whom it was intended.

### **Adaptability**

The group identified two main components to adaptability: cognitive-behavioral and emotional. The first of these, the cognitive-behavioral component, has to do with an individual's rational response to changes and considers factors relating to an individual's sense of control, situational awareness, stress and emotion management, and flexibility in analytical processes. By contrast, the emotional component deals with an individual's unconscious reactions to change and that individual's baseline emotional state (Perez-Fuentes et al., 2020). Given the target demographic of marginalized groups, who disproportionately experience emotionally traumatic events, as well as the intent of capturing adaptability as a conscious skill, the cognitive-behavioral component was determined to be more applicable to the intended application with the target demographics of Intrare. Thus, more questions related to cognitive-behavioral responses to change were selected and developed than questions related to emotional responses. However, in the interest of following the suggestions of established research on adaptability, some questions related to the emotional component were compiled while avoiding those which would punish users for their undue emotional trauma.

### **Communication**

Based on a review of soft skill assessments that measured the communication skill, the team created an initial list of questions. For example, the Communicative Competence Scale (CCS) assesses users on five dimensions of communication, including general competence, empathy affiliation/support, behavioral flexibility, and social relaxation (Rubin et al., 2020). Development Partnership focuses on providing solutions for performance-critical areas such as leadership, talent

management, and communication. Their communication assessment split questions into categories such as words, voice, and body language (Development Partnership | Delivering Management Consultancy Solutions That Work, 2022).

While the target demographic of marginalized groups may in some cases be disadvantaged due to differing mother languages, these concerns were found less relevant due to the fact that jobs offered through Intrare are Spanish-speaking and due to potential advantages stemming from multilingualism. With an understanding of the target components required to evaluate this skill, the related questions were compiled and developed.

### **Diligence**

In researching psychological definitions of diligence as defined by the team, the team found psychological ownership had particular relevance and intersection with the team's definition of diligence. Five key components were identified from research on psychological ownership, these being accountability, self-efficacy, self-identity, sense of belongingness, and territoriality. Accountability is both an individual's expectations of self and others, and an individual's responsibility to hold those groups and themselves to those expectations. Self-efficacy concerns an individual's self-confidence in their capabilities and capacity for success. Self-identity and sense of belonging concern an individual's conception and ownership of self and of group. Lastly, territoriality is posited as a fifth component of psychological ownership relating to an individual's dominion over physical, intellectual, and social objects and concepts (Avey et al., 2009). Of these, sense of belonging was identified as potentially problematic due to the definition of marginalization and Intrare's user base of marginalized groups, while the other four were deemed appropriate. These remaining four formed the basis of the questions developed and compiled for use.

## **Intellection**

While researching intellection, the team found this skill to be organized into two categories, critical thinking and intellectual curiosity. Critical thinking can be further broken down into reasoning, analyzing, problem solving, reading comprehension, scientific thinking, and creative thinking (Sarigoz, 2012). When the team looked into intellectual curiosity, we utilized the five dimensional curiosity scale (5DC), which was developed by a team in the psychology department at George Mason University. The 5DC is broadly composed of exploration, stress tolerance, and social curiosity (Kashdan et al., 2020). The 5-point Likert and the 5DC helped the team compile a list of questions from both sources.

## **Organization**

The team started by looking at the NADINE assessment's sub-skills of organization: work efficiency, time management, independent work, and generalized organization (Tountopoulou et al., 2020). The questions within the NADINE assessment are private, so the breakdown helps guide further research and potential questions. One of these assessments included ten components of effective time management behavior developed by literature review, focus groups, and expert insight (White et al., 2013). These components provided more guidance on how to measure these aspects. Additionally, more general job performance assessments with accessible questions pertaining directly to organization and connecting to the previously listed sub-skills such as work efficiency, planning, and time management were used in development of the questions (Ramos-Villagrasa et al., 2019).

Based on these resources, we drafted questions on organization by combining direct questions and the critical concepts underlying organization. However, due to the specific nature of our assessment, many aspects needed to be rethought and rephrased. For example, the ability to

adequately track tasks is a sign of high organization, commonly done using a planner. However, we recognized that some users may not have access to this resource, which could negatively impact their scores. Another consideration was that some pre-employment assessments utilized workplace-oriented wording, yet not all users may have held traditional jobs. More generalized wordings are more accessible to these users and were thus implemented.

## **Teamwork**

The team began by reviewing four journal articles about teamwork assessment. Three of the four publications contained assessments with questions that could serve as direct examples (Strom & Strom, 2011; Loughry et al., 2007; Romero-Díaz de la Guardia et al., 2022). The fourth publication was used to categorize the questions from the prior three assessments (Varela & Mead, 2018). These categories included getting along in the team, contributing to the team's work, expecting quality, keeping the team on track, and thinking critically and creatively. These questions were further organized to determine which categories were the most important. The team attempted to maintain a representative ratio for each category based on importance for the custom-made assessment. The questions were narrowed down based on which questions most accurately assessed the sub-skill.

### **4.1.2 Simplification and Review**

After compiling a working draft of the questionnaire, the team underwent multiple iterations of revisions. First, we worked with Professor Jed Lindholm of WPI, who specializes in various elements of human resources, including performance management, HR analytics, and competency identification. He has numerous experiences with pre-employment assessments and provided advice on how we could improve our own. Professor Lindholm expressed that the questionnaire was similar to industry standards but suggested rephrasing some questions to

improve clarity. He also assisted in ranking the questions by importance to ensure that we evaluate the most critical aspects of each skill.

Next, we brought the revised draft to our sponsor for review. The main concern during this iteration was the higher-level language that may be too confusing or intimidating for users. Our group altered many of the questions in the assessment to increase its accessibility while maintaining the content. Next, we sent the questions to be reviewed by WPI's Office of Diversity, Inclusion, and Multicultural Education (ODIME) to ensure that the questionnaire wasn't introducing any unintended bias. Upon receiving approval from ODIME, we used Chat GPT to simplify the wording of each question. This approach allowed us to simplify some additional questions, but we closely examined each simplification to ensure that the original meaning of the question was preserved.

### **4.1.3 Milestones**

A major design consideration of the questionnaire was the engagement and accuracy of the assessment. To help increase both of these we implemented a set of milestones to allow users frequent progress updates and opportunities to take a break. In the case of the questionnaire, the milestone will be treated as a progress marker for users. Specifically, the team implemented three milestones that included 18, 12, and 12 questions, respectively. The lengths of each milestone ensured users were presented with three items per skill assessed in the first milestone and two items per skill in subsequent milestones. Utilizing this approach allows for the completion of seven items per skill through an engaging element for users as confirmed during multiple conversations with the team and Intrare, while also offering several opportunities for users to take a break if necessary. Shown below are how the three milestones look while taking the questionnaire.



*Figure 6. Milestones of Questionnaire*

#### **4.1.4 Order**

After drafting a set of questions, the group needed to determine the order in which they would be presented to the user. To aid in this process, the list of questions was first prioritized according to relevance to the target definition of each skill. With this prioritization set, two options were considered when determining the order in which questionnaire items would appear to the user: randomization and scrambling. Randomization in the case of the assessment would involve presenting questions from each skill at random each time the questionnaire is taken, while scrambling would randomize the questions once and present the items in this order every time the questionnaire is conducted. The scrambling approach was used to reduce scoring variability and bias of the questionnaire. More specifically, we used segmented scrambling by scrambling the set of prioritized questions pertaining to each milestone. Using a pseudorandom number generator from one to 18, the first milestone question order was chosen. The order for milestones two and three was decided with the same method and an upper limit on the pseudo-random number generator of 12.

#### **4.1.5 Translation**

While Intrare's user base is diverse, it is predominantly hispanophone. We translated the questionnaire into Spanish to make it accessible to the majority of users. This translation was done

using a fluent and native Spanish speaker at WPI. With a native speaker, we were able to keep the intent of the script and questions. The original annotated English version of the questionnaire will be preserved for future translations into other languages.

## **4.2 Front End**

A primary goal of this project was developing a front-end interface that allowed the users to take the assessment within the Intrare platform. We first developed a wireframe using Figma, a popular prototyping software. The wireframe was later built into a React App to allow for user testing and easy integration into the Intrare code base. The development process for the front-end included prototyping, heuristic testing, development, and user testing.

### **4.1.1 Prototyping**

The questionnaire was designed prior to the prototype being designed. Because of this, the interface was designed to make the process of taking the questionnaire as intuitive as possible. The initial prototype was designed to have two screens: an introduction page and a question page.

The prototype was presented to members of the Intrare team as well as the WPI advisory team for feedback. From these feedback sessions, the wording of the introduction page was changed, which is reflected in the final version of the front end found in Figure 19. Additionally, both the advisory team and the Intrare team suggested adding words of encouragement throughout the assessment to encourage users to complete as many questions as possible. One of the considerations for the questionnaire is that users may be intimidated by the number of questions and not finish answering which the encouragements are meant to combat. These were also implemented in the final version of the front end.

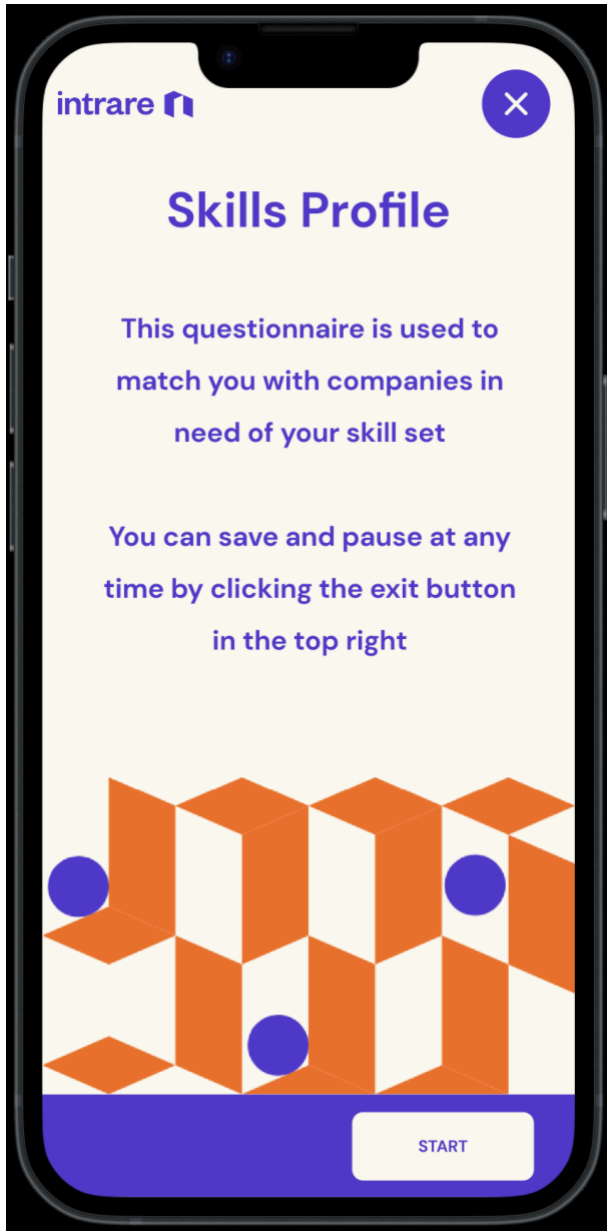


Figure 7. Wireframe Introduction Page



Figure 8. Wireframe Question Page

#### 4.1.2 Creating a Front End with ReactJS

After getting feedback on the prototype, the team moved on to developing the front end using ReactJS. The initial version of the app looked identical to the prototype and was intended to be used for Heuristic testing and A/B testing.



However, as the team further discussed ways to encourage users to answer more questions, changes were made. Through entire team discussions, it was decided that the questions would be split into three milestones so that the number of questions was less daunting. The status bar was therefore replaced with three circles, each representing a milestone of the questionnaire, that fill like a pie chart as each question is answered (Figure 11). Additionally, pop-up encouragements were added to motivate the user to progress through the questionnaire (Figure 12). The layout was slightly adjusted to account for these changes. The main changes were switching the Intrare logo to the icon, and removing the title from the questions.

Specific elements were added for the A/B testing, which is explained further in section 5.3.1.



Figure 9. Circles Representing User Progress



Figure 10. Pop-up Encouragement

## 4.3 Back End

The backend of the app has two components, a proposed database schema and a Javascript class. The Javascript, part of the ReactJS app, is intended to represent the user and their data. The user class stores, and contains methods for updating, the following information: current question, selected answers, and current subscores. This same information is stored in the proposed database schema (Figure 13). This database was not integrated into the platform by the team because of concerns of user privacy. However, Intrare was given the database schema and will implement it on their end.

The scoring is calculated in the Javascript user class, which is updated after the user answers each question; however, the score could alternatively be calculated in the database. For testing purposes, the team used the Javascript methods, but Intrare is able to easily change this at a later date. The scores are calculated by adding all of the answers together for each skill, which was originally proposed in section 3.2.4.

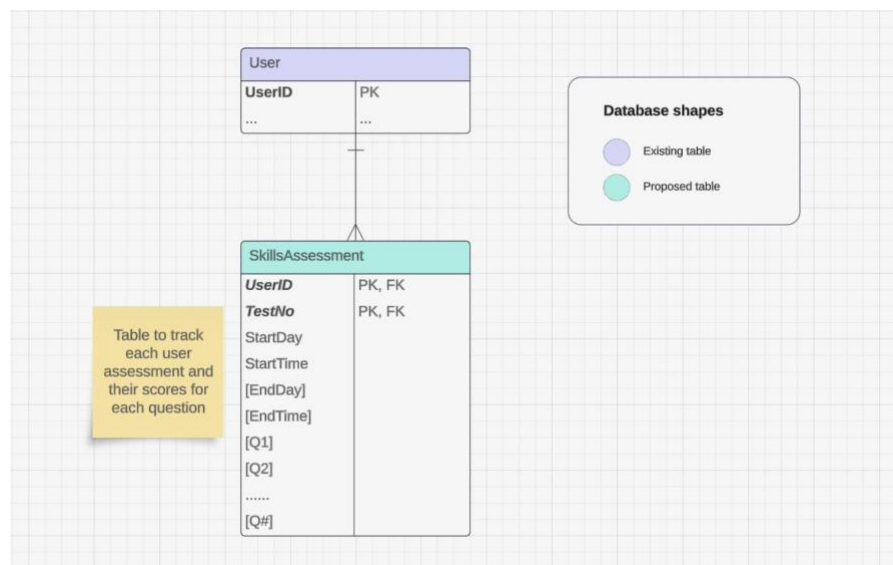


Figure 11. Database Schema

## 4.4 Skills Scoring Visualization

The team designed a custom scoring visualization to complement the pre-employment assessment and present the users with their results. The visualization is intended to give users a sense of pride in their skills and motivate them to improve their weaker skills. A main consideration throughout the design process was the need to present the scores positively and constructively. This pushed the team to combine the scoring systems of similar contexts with a custom-tailored interface. The scoring vector also needed to mesh well with the design of the questionnaire and be visually appealing to employers and users.

There are various ways to display the results of an assessment, but a common method is a scoring vector, especially in similar contexts (Fareri et al., 2021; Altomari et al., 2023). The team first considered a matrix score to clearly present the user's score for each skill and their overall performance. However, this scoring visualization seemed too complex for the user group, so we explored alternative examples to match the industry standard.

One example of a skills visualization is SkillNER, which uses data-driven techniques to extract soft skills from text. This tool attempts to address the difficulty of measuring soft skills quantitatively. SkillNER uses computational linguistics to pull and classify entities based on the O\*NET framework (Fareri et al., 2021). While their visualization matrix displays the soft skills needed for a specific job profile, it could apply to various contexts that require portraying the connections and clustering of different skills. Another soft skills assessment used a tree graph to identify patterns and interactions within a game interface. Each tree-graph node serves as a scoring vector that can be combined to create a final vector, and the vectors factor into a formula to obtain their score. The numeric results and the vectors work together to visually display and describe the

results to users (Altomari et al., 2023). Building off these instances, the team intended to use scoring vectors to capture the user's results and assist in the visual display.

The skills visualization went through multiple iterations with various deliberations over the design and intuitiveness. Based on earlier considerations, the matrix form seemed too complex, leading the shift to a more linear-based visualization. The first design utilized a pie chart to display the comparative scores amongst each skill. However, while this intended to highlight where the user did better or worse, it did not present where the user missed points and could be misleading to some users. While the results should encourage the user, it became clear that an overly simplified score would not provide any areas for improvement, so the team sought an iterative solution that addressed this concern. We also wanted to pair the overall score with a more detailed sub-skill breakdown for each skill category. While the team considered multiple designs, it was eventually determined that a Likert scale would be easy to implement based on the scoring system and would assist in providing a representative score.

Upon the first draft of the scoring visual, the team reached out to Professor Lane Harrison, who teaches at WPI and specializes in data visualization. After hearing more about the specific needs of our project and the concerns on the first draft, he suggested playing around with the design since it can be iterative and not linear progress. He also suggested implementing a red-to-green color scale to highlight these strengths and weaknesses, which we decided to add to the Likert scale to provide an intuitive way to read the results.

After further iterations, the team developed a new design, hoping to keep the strengths from past drafts and solve some of the previous concerns. The team moved forward with the pie chart but added in the missing points and the unanswered questions, which would also serve as a form of uncertainty. Rather than having one pie chart, there would be one for each of the six skills that

could be accompanied by the drop-down to display the sub-score skills through the color-coded Likert scale. However, with respect to time and the group's skill set in data visualization, we sought to create a clean integration into the assessment and prioritized the development of the pie charts (Figure 14).



Figure 12. Iterative Draft of the Scoring Vector

In these later drafts, we continued to brainstorm ways to keep the display understandable and intuitive on a mobile phone, which is the device utilized by the majority of users. We also needed to consider the compatibility in design between the questionnaire and results, pushing the team to utilize more of the Intrare colors and minimize the size of the results to include all on one page. Additionally, we wanted the design to be visually appealing enough to be presented directly to employers or have users post their results to their social media and display their results. The final version of the scoring prototype is shown below in Figure 15 and Figure 16.

Ultimately, due to the time constraints of the project and the team’s little experience with coding data visualizations, we worked to develop a detailed Figma prototype that would capture the proposed design above and add functionality. The prototype will later be developed and integrated into the assessment by Intrare employees.



Figure 14. Scoring Prototype

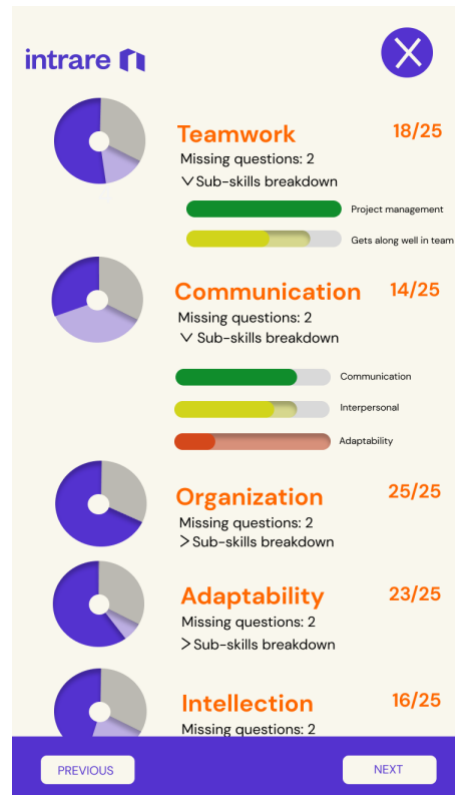


Figure 13. Scoring Prototype with Sub-skills

## 4.5 User Testing

User testing is imperative to a company's ability to meet the needs of its audiences. A/B testing is a technique used by larger corporations to understand the relationship between design elements and user experience (Fabijan et al., 2018). A/B testing involves randomly splitting subjects between a control (A) and test version (B) to compare which is preferred according to measured user activity. While A/B testing on a smaller scale does not produce enough responses

to provide statistically significant evidence, when paired with other forms of user feedback it can still provide some context into the user's perspective of the platform (Paulsson et al., 2022). With these considerations, our sponsor encouraged us to go forward with A/B testing. In this section, we will further develop the protocol.

The goal of our A/B testing is to understand how the users interact with the design of the platform and how specific elements impact the user experience. We specifically considered the design goal of accessibility for our target groups as inspiration for areas of improvement and potential problems. These included implementing a more black/white design for color blindness, increasing the font size for readability, adding an auto-update scroll, and testing different ways of tracking progress. However, due to the time constraints of the project and the limited user base, we decided to let Intrare decide on two areas for testing. While we could not implement all test versions, this approach provides feedback and potential suggestions for future improvements. The two sub-versions tested include increasing the font size to increase readability and adding pictures that aid the understanding of the Likert scale to mitigate potential language barriers.

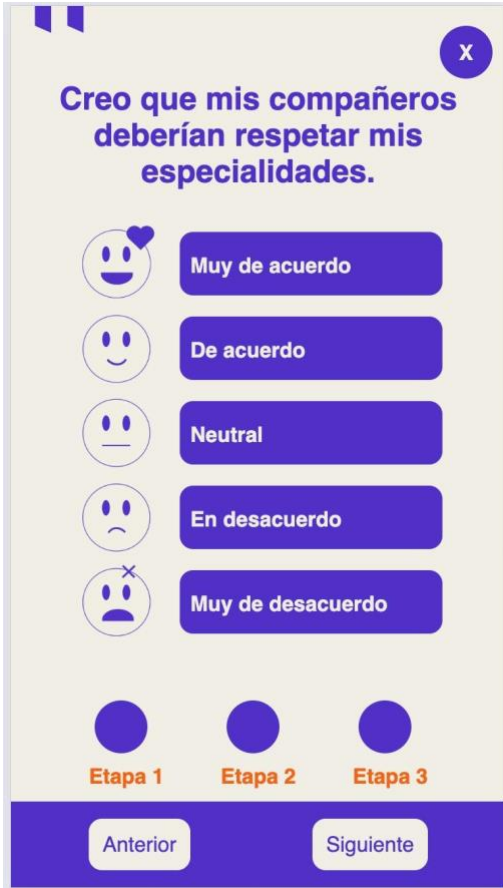


Figure 15. Language Barrier B Version

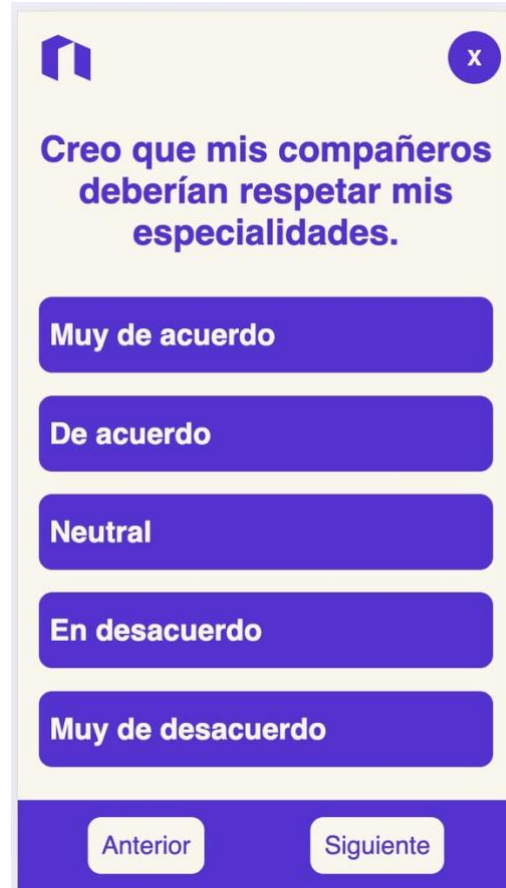


Figure 16. Font Size B Version

## 4.5.2 Constraints

There are a few limitations that should be considered when analyzing the results from A/B testing. One consideration is the limited sample size since Intrare is a smaller company, which makes it difficult to obtain statistically significant findings. However, there is an accompanying smaller user base, which likely does not require large-scale user testing. Yet, with user testing on this scale, we are also limited in recognizing potential small treatment effects that may have more intense impacts on a broader scale. The results should also evaluate the potential for confounding effects, especially with the limited information from users (Bhat et al., 2020).



## **5. Results**

After the design and execution of the project, the results are discussed in the following sections. First, we will present the front end results and walk through the progression of the assessment. Then, the user feedback results will be discussed including heuristic and user testing. The results included will provide both areas of strengths and weaknesses, which can be iterated on in the future to better the assessment.

### **5.1 Front End Results**

The development of the ReactJS App created an interface that was viable for testing. The app can be further iterated upon by Intrare, but the current version is viable for initial use by users. This section breaks down the interface by element, diving into the function of each element. Translations for each page and the full questionnaire are available in Appendix D .

#### **5.1.1 Introduction**

The first page that the user sees upon opening the questionnaire is the Introduction page (Figure 19). This page encourages honesty when answering the questionnaire as well as provides an estimate of the time commitment required of the user. The user is able to start the questionnaire with the button on the bottom of the screen.

#### **5.1.2 Questions**

Upon starting the questionnaire, the user is presented with the screen seen in Figure 20, which is the layout for all 42 questions. At the top of the screen is an exit button which brings the user back to the Introduction page. Below that is the question, which changes as the user navigates the questionnaire. Underneath the question is a five point Likert scale; the same scale is shown for

each question. Below that are the status circles which represent each stage of the questionnaire. As the user progresses the circles slowly fill as a pie chart (Figure 21). To encourage the users throughout the questionnaire, encouragements appear for 3 seconds (Figure 21). The pop-ups are evenly spaced amongst the 42 questions, occurring every four questions on average. However, since specific encouragements must occur at specific points in the questionnaire, such as those for completing a stage, there is variation in how often the pop-ups appear. A full list of encouragements as well as when they appear is available in Appendix E.

### **5.1.3 Conclusion Page**

After the user completes the questionnaire, they are brought to the conclusion page (Figure 22). The text congratulates the user on completing the questionnaire and directs them to a separate page to view their results.



Figure 17. Questionnaire Introduction Page

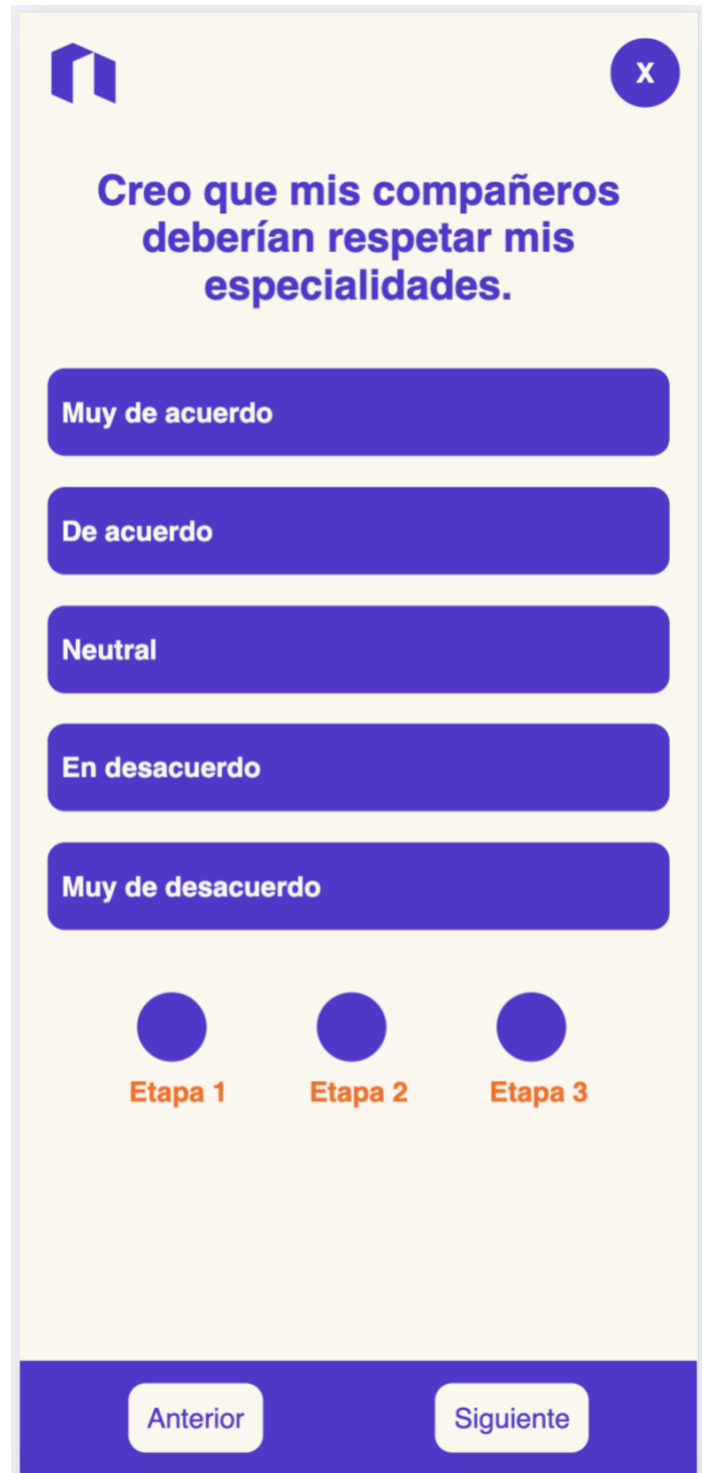


Figure 18. Questionnaire Question Page

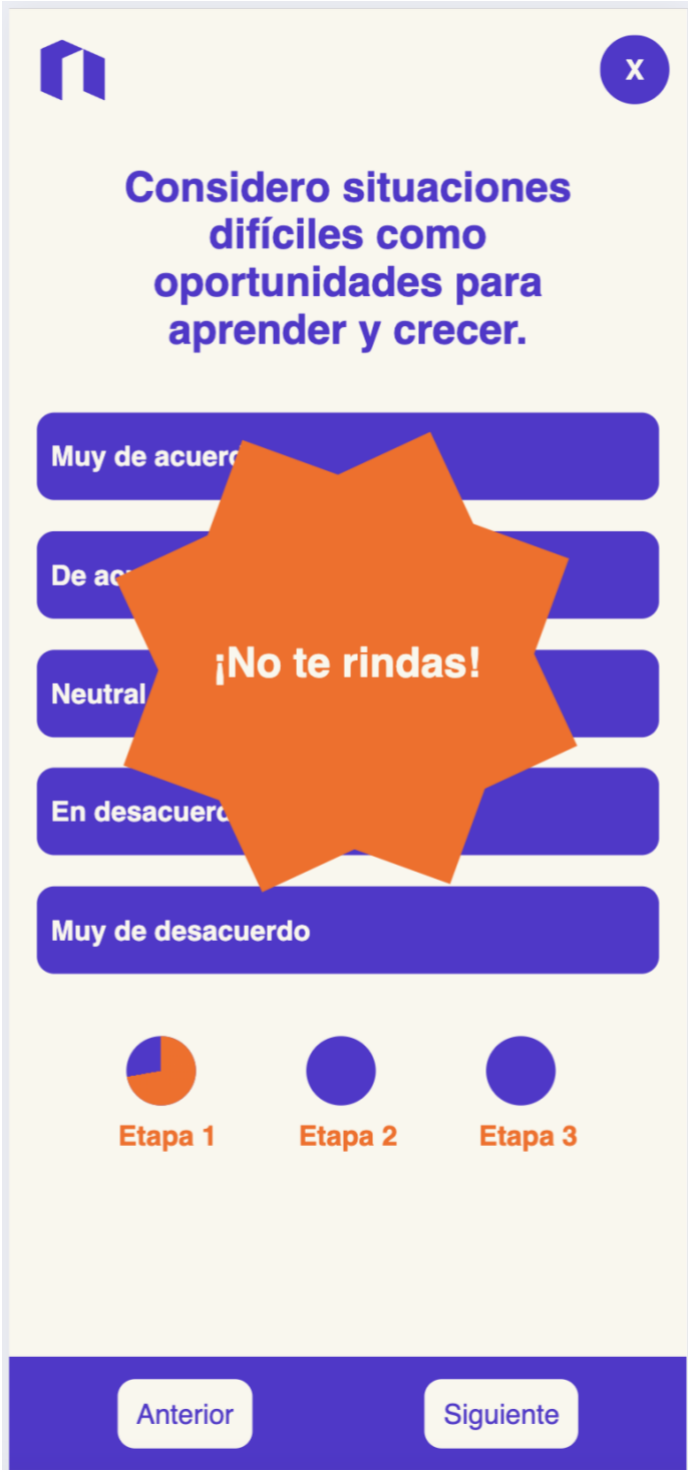


Figure 19. Question Page with Encouragement

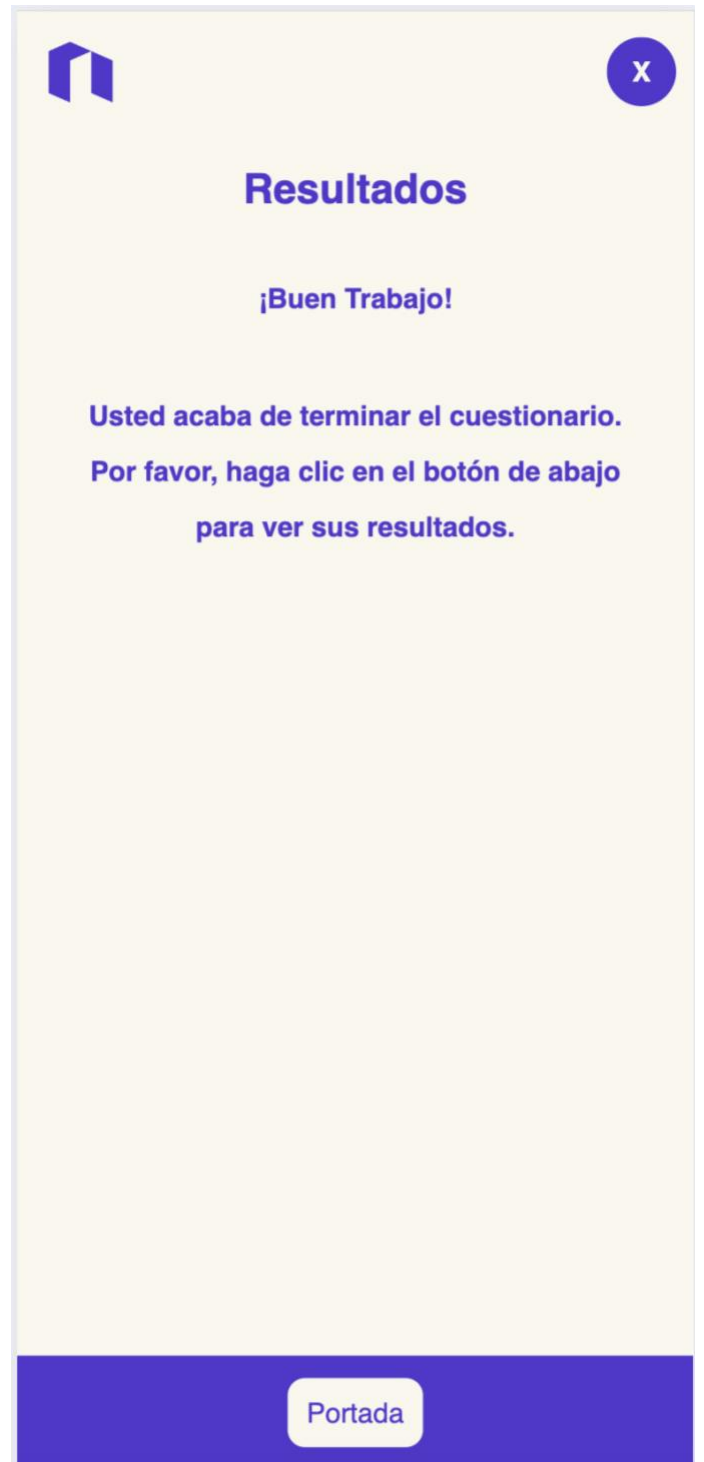


Figure 20. Questionnaire Conclusion Page

## **5.2 Heuristics**

Part of the user testing the team conducted was an heuristic evaluation with Intrare employees. In the following section, we will discuss both the protocol we developed, feedback we received, and the corresponding changes we made to the assessment.

### **5.2.1 Protocol**

The team used usability heuristics to evaluate the interface of the assessment from subject matter experts, which in this case were Intrare employees. The severity scale and heuristics were compiled into a document for the employees to rank and comment on the heuristics. The evaluation document was attached to the protocol, found in Appendix A, and sent to Intrare employees to provide clear instructions.

### **5.2.2 Results**

After sending the heuristic document to Intrare, the evaluation was conducted over the course of two weeks. The compiled list was sent back to the team with comments and an average severity score for each heuristic.

The majority of the heuristics measured received a severity rating of zero or within 0.5 points. This signifies that there were no alarming or concerning areas of the assessment. However, three sections received a score of one, and these elements served as an area that required further investigation. The table below is a condensed version with sections of concern (Table 6). Section 1 received a score of 1.2, and the comments regarded the wording of questions as difficult to understand, which the team further evaluated and changed. For example, the question “I have a way to maintain my schedule” was developed to “I have a way to keep track of my schedule.” This rewording was intended to increase the question’s clarity and accessibility for all users. Looking

into section 2, which received a score of 1, these comments suggested moving the progress bar to the top of the screen. From here, the team created two versions to see how the difference in location would impact user experience. After creating the mock-ups and through further discussions, we decided to keep the progress bar at the bottom to avoid conflict with the other design elements at the top of the page, including the Intrare logo and exit button. Finally, section 8, which scored a 1, suggested adding an alert that would warn the users if they attempted to advance without responding to the current question. This feature was added to the assessment's next iteration and can be seen in the earlier images of the assessment. With the initial interface testing complete, we then moved into the A/B testing phase of the project.

*Table 6. Partial Heuristics Results*

<b>Heuristics</b>	<b>Description</b>	<b>Severity Rating (0-4)</b>	<b>Comments</b>
1. There is clear language and grammar within questions	Text should be grammatically correct and make sense.	1.2	The language is understandable, but there are two questions (I have a way to maintain my schedule, I can change my way of speaking depending on the situation) that I found difficult to understand. I feel that they could be reformulated in a better way. / The question "Can I change..." is repeated twice.
2. Progress through the test is visible and understandable	Ensure that the number of questions answered and remaining is obvious and time remaining is clear.	1	Given that there is no mandatory time limit for completing the test, there is no detail on this point. I would like the progress bar to be at the top.
8. Feedback is appropriate	System feedback should be clear about what action is required.	1	Adding a tooltip or alert for attempting to advance without responding or abandoning the questionnaire would provide additional assistance in understanding the actions.

## **5.3 Preliminary User Results**

The team planned to conduct A/B testing as the primary form of user testing. However, Intrare could not complete the full version of A/B testing within the project's time constraints. Yet, there was some preliminary testing to evaluate basic metrics regarding user interaction, including the demographic breakdown and the assessment completion rate. The team provided Intrare with a complete A/B testing protocol and sub-test versions for future testing.

### **5.3.1 Protocol**

With access to more of the platform, Intrare executed the testing and provided the results for analysis. The testing tracked two main features: the user's ability to finish the assessment and the efficiency of the test based on the user's results. The testing was conducted from January 4th, 2024 through February 7th, 2024, with 75 Intrare users.

### **5.3.2 Results**

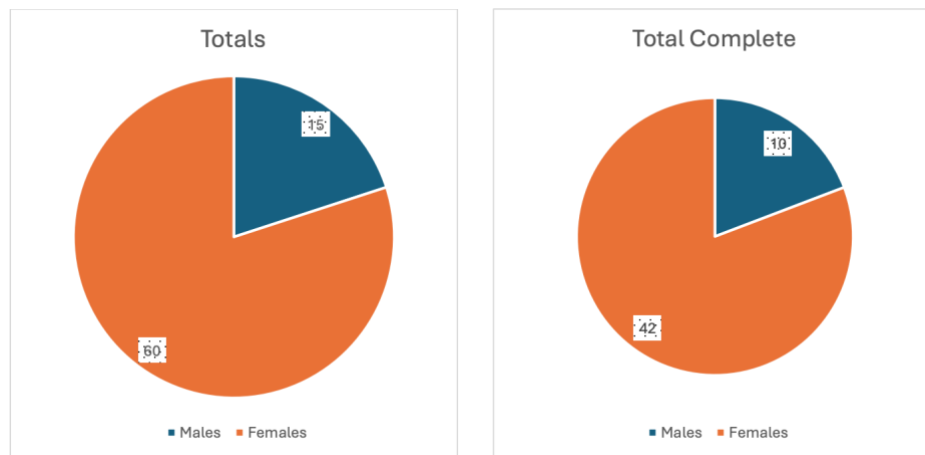
The following section will outline the results we obtained from Intrare from the preliminary testing, including how many users completed the assessment, the time to completion, and the impact on job results.

Of the 75 users who took the assessment, 60 were female, and 15 were male. The test was shown randomly to users, and the majority female audience could be more reflective of the Intrare user base. However, since there is a discrepancy in the gender ratio, there may need to be additional user testing to ensure there is no gender bias.

A vital project goal was to maximize the number of users who finished the assessment. 69.3% of the users, or 52 individuals, completed the questionnaire. However, it is unclear where some users quit the test. The team implemented a milestone system to help reduce the time

requirement and to serve as a natural break in the questionnaire. There should be further research into what stage users are halting the assessment to determine if there are any trends in behavior. For example, if users quit early on, this may be linked more to the difficulty of the assessment. However, if users abandon it later, this may indicate that the questionnaire is too long or unengaging.

Further investigating the gender distinctions, 70% of females completed the assessment (42 individuals), as compared to just 66% of the males (10 individuals). Again, this discrepancy requires additional analysis to ensure that there is no gender bias present. However, this difference may reflect general interactions of the Intrare user base based on gender. Regardless, this could be a potential area of improvement for both the assessment and the general platform.

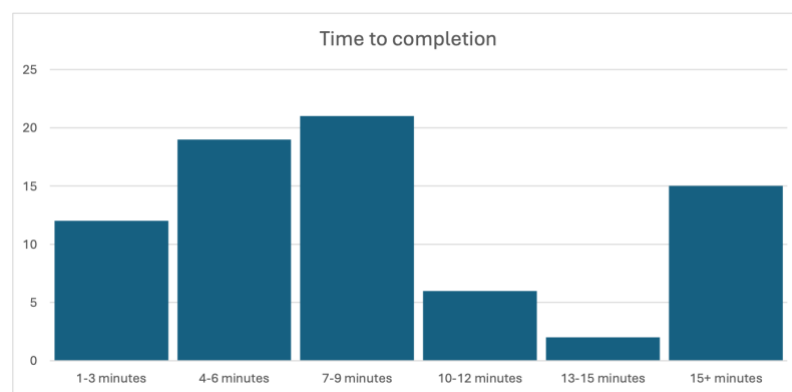


*Figure 21. Assessment Demographics by Gender*

Another result Intrare provided was the time users took to complete the assessment. While there is no gender distinction in this result, it provides more general information about the total time needed to complete the questionnaire. This metric is a critical result, as a key consideration throughout the project was to reduce the time requirement of the assessment. The team planned for the questionnaire to take around 15 minutes or less. The provided result captures the total time spent, and the individuals who took longer than 15 minutes represent those who didn't complete



the questionnaire. As shown in Figure 24, most users took between four and nine minutes to complete the assessment. This shows the team was successful in keeping the time commitment low to respect the user's time. However, we want to ensure users take their time on the assessment. The team designed a shorter questionnaire, but it would be unlikely for the users to receive accurate results from such a short questionnaire (1-3 minutes). More analysis could be conducted to explore the user's results and ensure that user's are properly filling out the assessment, rather than clicking the answers they presume to be correct.



*Figure 22.* User Time Breakdown on Assessment

Another result is the user breakdown by demographic, which connects to the underlying goal of creating an accessible assessment. Similar to the gender breakdown, the randomized test likely aligns more with the demographics of Intrare's users. The results appeared to capture all of the groups Intrare serves. Yet, by comparing the results to more standard metrics on Intrare's user base, there will be assurance that there is feedback from all groups. If there were any discrepancies, it would be imperative to conduct more research to ensure that all groups could complete and understand the assessment.

Additionally, there is no differentiation between the users who did and did not finish. This missing element should be investigated to understand how different demographic groups interacted with the assessment and limit potential biases. For example, only three individuals who identified

as refugees took the questionnaire. However, it is unclear what percentage of them were able to complete the assessment.

Another area to investigate further would be how different demographic groups scored concerning each skill. While all users have unique experiences, there may be a correlation between certain groups and skills.

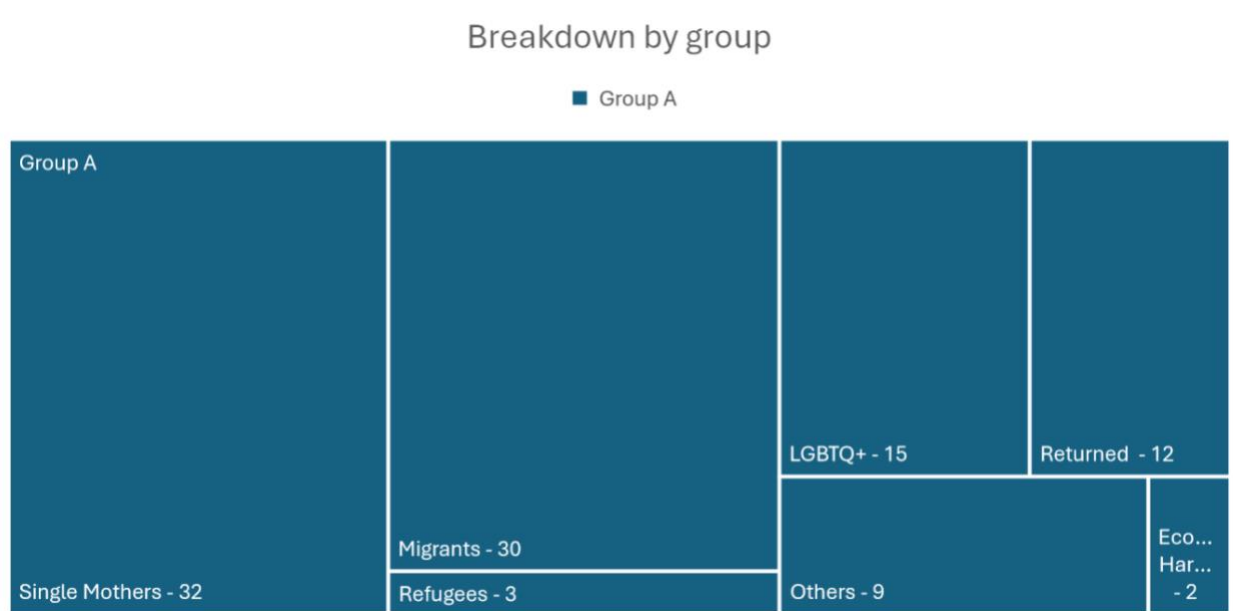


Figure 23. User Testing Demographics

Lastly, both the average number of jobs presented and the number of jobs users replied to were higher for the test group that took the assessment compared to a similar-sized sample of users. While this was a preliminary test, it appears that the soft skills assessment strengthened the user profile resulting in more matches. Additionally, there was a higher rate of response to recommended jobs which could mean that the users also felt these job matches were a better fit for themselves. The team sought to strengthen the user profiles on the Intrare platform and increase job matches. These appear to be accomplished during the user testing but should be continued to be monitored as the assessment is fully integrated into the website.

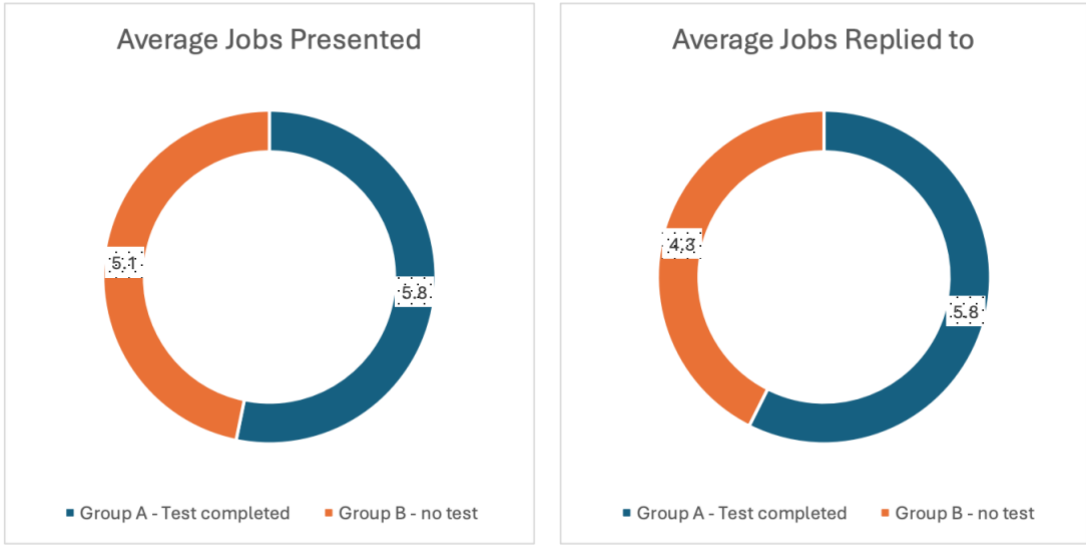


Figure 24. Assessment Impact on Job Results

## **6. Recommendations & Conclusion**

Given the limited timeframe of this project, some elements of the final product and its development were unable to be completed by the team. That said, these components are essential to the development and improvement of an effective skills consideration and assessment in the context of many-to-many job matching, and they are thus presented in the following section as recommendations for future work.

### **6.1 Focus Group**

The first recommendation would be to administer a focus group with Intrare users concerning their thoughts on both the skills assessment and the platform on which it is administered. This would aid the validation and potential modification of the assessment greatly for several reasons. Specifically, it would supplement existing testing data to enable meaningful improvement to the questionnaire's platform; provide meaningful insight to the accessibility, relevance, and accuracy of the questionnaire itself with regards to the intended user base and the ultimate application in a many-to-many job matching integer program; and serve more generally to help understand and address the gap between Intrare's employees and its target demographic.

The first of these reasons has to do with the nature of the testing that has thus far been conducted. While heuristic and A/B testing have their merits, they do not produce a comprehensive understanding of how users interact with a platform on their own because they do not allow for open-ended, comprehensive user feedback. As a result, it is not possible to directly implement meaningful changes to the original prototype in response to user preferences and habits. Supplementing this understanding of user activity is paramount to ensuring that the assessment suits the users it is intended to serve.

Additionally, conducting a focus group would serve to validate the questionnaire itself. Great attention was paid to the development of questions that would theoretically be both accessible to the user and relevant to the skills to be assessed. The methodology implemented to develop the questionnaire was dedicated heavily to identifying and developing research-backed questions for the target skills and adapting these questions to fit the target audience of refugees, migrants, vulnerable women, single mothers, members of the LGBTQI+ community, and returned peoples. However, having invested this time and effort into the theoretical nature of the questionnaire, there is, as of yet, no confirmation that the theoretical has manifested in the actual product. To this end, user feedback would help to determine whether the questions themselves achieved the objectives of being both relevant to target skills and accessible to Intrare's target demographic. Ensuring both of these objectives were met is essential due to the nature of the analytical methods to which the results of this questionnaire are to be applied. Analytics, including many-to-many integer programs such as the one currently in use at Intrare, require the use of reliable, accurate data to provide meaningful, accurate results. Consequently, inaccurate, unreliable data result in exacerbated inaccuracies and unreliability in the results provided by analytical methods. This means inaccuracies and irrelevancies produced by the questionnaire through its questions and accessibility will yield results in the downstream analytical methods that at best will be meaningless and at worst will actively harm the already vulnerable groups Intrare intends to help.

Lastly, a focus group will serve a more general purpose of bridging the gap between Intrare and its user base. In working on this project, the group noticed a gap between the group's perspective and that of the vulnerable populations for which the group was working. This sentiment was echoed in correspondence between the group and Intrare. This gap can best be

addressed through communication and interaction between the developers of Intrare's user products and the users themselves. This sort of communication is intrinsic to a focus group, and thus a focus group would help to facilitate greater understanding of the user base's need and ultimately result in a better and more useful user experience.

## **6.2 Observe and Adapt to Trends in Data**

While extensive effort has been made to ensure that the questionnaire encourages honest responses from its users and in turn produces accurate, useful information pertaining to user skills, it is likely that the questionnaire will in some cases fail to encourage honesty, distinguish users from each other, or both simultaneously. With this in mind, it is recommended that Intrare monitor the trends in user responses to the questionnaire and adapt it accordingly. This adaptation can come in many forms based on the particulars of the user data, but two particular ways in which the questionnaire can be adapted are by assigning questions different scoring weights based on similarities in user responses and rephrasing questions based on their propensity to encourage specific answer choices. The former approach serves to produce more meaningful differences in user scores to better distinguish users from each other, while the latter approach would serve to encourage more honest introspection and responses to the questions.

Based on user answers, certain questions can be weighted to have greater or lesser effect on the final skill score. Such an approach is proposed by Chakrabartty to address the limitations of the traditional five point Likert scale. To account for respondent biases such as Central Tendency, Acquiescanal, and Social Desirability, as well as to account for the variability of perception regarding what it means to "Strongly Agree" with a question and how "Strongly Agreeing" differs from simply "Agreeing" or even "Disagreeing," approaches are proposed weighting questionnaire items according to frequency of certain answers. By doing this, each

individual item within the test questionnaire could be found independent, allowing for usual analysis of the results. Additionally, these approaches resulted in a more reliable questionnaire (Chakrabartty, 2014). While, at the time of writing, little user response data exists with which to analyze and perhaps implement such a solution, it is recommended that such analysis be done and an approach similar to those proposed be implemented should user data indicate it is appropriate.

Additionally, certain items in the assessment may require revision according to trends in user data. While existing biases in Likert scale assessments may explain certain trends as previously mentioned, it may prove prudent to consider whether trends in response data are related to the original composition of the individual questions in the assessment. Specifically, certain items may seem to encourage certain responses regardless of verity, which would in turn render results garnered from these responses less accurate. Without substantial answer data it is difficult to make specific recommendations, however one example of an item that could fit this description is “In group settings, I care about doing things well.” This phrasing could be particularly vulnerable to Social Desirability bias and may also fail to engage the respondent in self-reflection, as it is generally expected that one should care about quality of work. However, the intended effect of measuring a respondent’s prioritization of work quality in a group setting is still important to the assessment of their teamwork skill. While weighting the question so as to reduce the effect of this item in the score of respondents would allow for better analysis on the final results, it may also have the effect of punishing users with greater Central Tendency bias, trading one problem for another. As such, it is recommended that such questions instead be rephrased to better provoke honest introspection and self-reporting.

## 7. Reflection

As the team concluded the project, we were able to reflect back on our experiences. The following section will include more details on the project evolution, constraints, learning outcomes, and collaborating as a team.

### 7.1 Project Evolution

Over the course of this project, the objective and methods in which that objective would be achieved varied significantly. This project began with a focus on several algorithmic methods by which to incorporate a skills component to an existing many-to-many job matching algorithm. The group considered algorithmic fairness as a potential component of the final deliverable. This deliverable was at several points considered to take the form of a game or set of games, but was eventually determined to be a questionnaire. The group also considered the inclusion of both hard and soft skills as part of the scope of the assessment.

Ultimately, most of these considerations were deemed to be unfit for the objective we were given. The aim of this project was to develop a method by which skills could be elicited from vulnerable groups such as migrants, refugees, single mothers, vulnerable women, the LGBTQI+ community, and returned peoples. All decisions regarding design considerations were made with this user base in mind, and through this lens topics such as fairness were deemed beyond the scope of the project. From here, certain solutions such as games and the inclusion of hard skills were deemed inaccessible to the target user base. This narrowed the focus of the group to a questionnaire-based assessment of six soft skills. A corresponding interface was developed, again with the consideration of accessibility at the forefront of the design.



## 7.2 Constraints

There were several constraints which impacted the progress of the project. One of the main constraints was the access to the users, since they are based in Mexico. The locational and language differences prohibited the team from directly interacting with the users and obtaining feedback on the assessment. Intrare was able to get some feedback from both Intrare employees and users. However, the ability to get more direct feedback would assist in ensuring that the assessment we created achieved our initial goals.

Another limitation was understanding the user base, since the demographics of the user base are very different than those of the team. Intrare serves several marginalized groups who all have different backgrounds and experiences. The team often had to reevaluate and put aside their own experiences to prioritize the useability of the assessment for the target audience. At times, one approach seemed optimal, but upon further thought may not best serve the users. The team often iterated on ideas, shifted focus, and improved the final product.

Similarly, the variety of users made it hard to account for all the unique needs and experiences. While the assessment needed to capture necessary information, it remained generalized to avoid favoring any particular group. For example, one consideration was that some users may not have held or had the opportunity to pursue traditional jobs in the past. The team kept all groups and their unique experience in mind throughout the development to achieve a fair assessment.

Time also largely impacted the team and the project. While we made consistent progress throughout the year, we found that due to the custom needs of the assessment, there was less time to get feedback from users as well as analyze the results. In the future, Intrare can obtain this

information and make changes as needed; however, the student team also wishes we could do this evolution over time ourselves.

## **7.3 Learning Outcomes**

During the progression of the project, the team was able to learn from their experience, each other, and the sponsors. The team's initial research looked into soft skills, user engagement, and algorithmic fairness. Additionally, user engagement and algorithmic fairness provided valuable insights into enhancing the user experience while ensuring ethical and unbiased solutions. These topics can potentially be applied to similar areas of work in the future.

The user base of Intrare and thus the project is vulnerable groups which require unique considerations. The team has had to consider vulnerable groups for previous academic projects but not to the rigor required of this project. During the entire research and development phases of the questionnaire, the needs and backgrounds of the users had to be considered. This was a fulfilling experience and provided insights into effectively addressing the needs of vulnerable groups.

Finally, the team was able to learn immensely from working with Intrare. This cooperation gave the team insights about the operations of a social startup. The team came to understand the constraints a social startup is under through the biweekly meetings and email correspondence with Intrare. While these constraints impacted development the team was able to work collaboratively within the team and Intrare to develop creative solutions and keep the project on track. They also learned the process of planning and leading team and project meetings. Our collaboration with Intrare provided insights into the operations of a social startup, creative problem solving and enhanced our project management skills.

## 7.4 Working as a Team

Due to the collaborative nature of this MQP, it was necessary to communicate frequently with all relevant individuals. The team communicated mainly over text, but also had regularly scheduled meeting times to work through more in depth challenges. With the Intrare team and MQP advisors, the team primarily communicated through email and meetings. This communication was critical to ensure that the team was on track and the work aligned with Intrare's vision.

The team also shared responsibility throughout the project in various ways. Team members took ownership of various aspects of the project depending on their expertise. The team rotated the who led meetings with advisors and sponsors. This team member would send the agenda to attendees prior to the meeting as well as keep the meeting on track. In the event that a question was asked and the lead wasn't able to answer it, they would direct the question to the team member who had ownership over that aspect of the project.

Through team meetings, deadlines, and feedback sessions, the team was able produce high quality work and remain on track. Team meetings occurred at least twice per week, but up to five times per week. These meetings were used to independently work, set deadlines, and review ideas and work. Deadlines were used to ensure that the team was on track to complete the project and receive feedback from the advisors and Intrare. Feedback sessions were also critical, so the team was able to provide the highest quality work.

## 8. References

- Ahani, N., Andersson, T., Martinello, A., Teytelboym, A., & Trapp, A. C. (2021). Placement Optimization in Refugee Resettlement. *Operations Research*, 69(5), 1468–1486. <https://doi.org/10.1287/opre.2020.2093>
- Al-Sa'di, A., Yamjal, P., Ahmad, E., Panjabi, R., McPhee, C. A., & Guler, O. (2023). Assessing Educators' Soft Skills: Developing a Self-Assessment Instrument. *Administrative Sciences*, 13(9), 208. <https://doi.org/10.3390/admsci13090208>
- Altomari, L., Altomari, N., & Iazzolino, G. (2023). Gamification and Soft Skills Assessment in the Development of a Serious Game: Design and Feasibility Pilot Study. *JMIR Serious Games*, 11, e45436. <https://doi.org/10.2196/45436>
- Analysis and Recommendations on Data Collection: Marginalized and Minority Groups Protection Cluster (March 2021) - Somalia*. (2021, August 9). ReliefWeb. <https://reliefweb.int/report/somalia/analysis-and-recommendations-data-collection-marginalized-and-minority-groups>
- Appelt, K. C., Milch, K. F., Handgraaf, M. J. J., & Weber, E. U. (2011). The Decision Making Individual Differences Inventory and guidelines for the study of individual differences in judgment and decision-making research. *Judgment and Decision Making*, 6(3), 252–262. <https://doi.org/10.1017/S1930297500001455>
- Aridi, M., Ahmad, M., Ghach, W., Charifeh, S., & Alwan, N. (2023). Development and Validation of the Soft Skills Questionnaire for Nurses. *SAGE Open Nursing*, 9, 23779608231159620. <https://doi.org/10.1177/23779608231159620>

- Bartram, D. (2007). Increasing Validity with Forced-Choice Criterion Measurement Formats. *International Journal of Selection and Assessment*, 15(3), 263–272.  
<https://doi.org/10.1111/j.1468-2389.2007.00386.x>
- Bertsimas, D., Farias, V. F., & Trichakis, N. (2011). The Price of Fairness. *Operations Research*, 59(1), 17–31. <https://doi.org/10.1287/opre.1100.0865>
- Bhat, N., Farias, V. F., Moallemi, C. C., & Sinha, D. (2020). Near-Optimal A-B Testing. *Management Science*, 66(10), 4477–4495. <https://doi.org/10.1287/mnsc.2019.3424>
- Bogen, M., & Rieke, A. (2018). *Help wanted: An examination of hiring algorithms, equity, and bias* (United States of America) [Report]. Upturn. <https://apo.org.au/node/210071>
- Chakrabartty, S. N. (2019). Scoring and analysis of Likert scale: few approaches. *Journal of Knowledge Management and Information Technology*, 1(2), 31-44.
- Chauhan, P. S., & Kshetri, N. (2022). The Role of Data and Artificial Intelligence in Driving Diversity, Equity, and Inclusion. *Computer*, 55(4), 88–93.  
<https://doi.org/10.1109/MC.2022.3149017>
- Chen, Y., Liu, Y., & Podimata, C. (2019, November 10). *Learning Strategy-Aware Linear Classifiers*. arXiv.Org. <https://arxiv.org/abs/1911.04004v4>
- Chirica, A. (2021). A Sociological Study of HR Professionals’ Discourses in Regards to the Digital Transformation of Human Resources Practices. *Revista Română de Sociologie*, 32(1/2), 39–55.
- CodeSignal. (2023). <https://app.codesignal.com/evaluation-result/mEENbpuFcwFZLvBEi>
- Development Partnership | Delivering management consultancy solutions that work. (2022). Retrieved January 30, 2024, from <https://the-dp.co.uk/>

- Díaz-Arancibia, J., Rusu, C., Pow-Sang, J., & Roncagliolo, S. (2013). A cultural-oriented usability heuristics proposal. *Association for Computing Machinery*.  
<https://doi.org/10.1145/2535597.2535615>
- Escolà-Gascón, Á., & Gallifa, J. (2022). How to measure soft skills in the educational context: Psychometric properties of the SKILLS-in-ONE questionnaire. *Studies in Educational Evaluation*, 74, 101155. <https://doi.org/10.1016/j.stueduc.2022.101155>
- Hermawati, S. & Lawson, G. (2016). Establishing usability heuristics for heuristics evaluation in a specific domain: Is there a consensus? *Applied Ergonomics*, 56, 34–51.  
<https://doi.org/10.1016/j.apergo.2015.11.016>
- Nielsen, J. (1994). *Usability Inspection Methods*. John Wiley & Sons.  
<https://www.nngroup.com/books/usability-inspection-methods/>
- Fabijan, A., Dmitriev, P., McFarland, C., Vermeer, L., Holmström Olsson, H., & Bosch, J. (2018). Experimentation growth: Evolving trustworthy A/B testing capabilities in online software companies. *Journal of Software: Evolution and Process*, 30(12), e2113.  
<https://doi.org/10.1002/smr.2113>
- Fareri, S., Melluso, N., Chiarello, F., & Fantoni, G. (2021). SkillNER: Mining and mapping soft skills from any text. *Expert Systems with Applications*, 184, 115544.  
<https://doi.org/10.1016/j.eswa.2021.115544>
- Finocchiaro, J., Maio, R., Monachou, F., Patro, G. K., Raghavan, M., Stoica, A.-A., & Tsirtsis, S. (2021). Bridging Machine Learning and Mechanism Design towards Algorithmic Fairness. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 489–503. <https://doi.org/10.1145/3442188.3445912>

- Giesecking, J. J. (2018). Size Matters to Lesbians, Too: Queer Feminist Interventions into the Scale of Big Data. *The Professional Geographer*, 70(1), 150–156.  
<https://doi.org/10.1080/00330124.2017.1326084>
- Green, A., Parke, C., Hoole, C., & Sevinc, D. (2021). Unlocking inclusive growth by linking micro assets to anchor institutions: The case of skilled overseas migrants and refugees and hospital jobs. *European Urban and Regional Studies*, 28(4), 450–465.  
<https://doi.org/10.1177/09697764211014231>
- Hane-Weijman, E. (2021). Skill matching and mismatching: Labour market trajectories of redundant manufacturing workers. *Geografiska Annaler: Series B, Human Geography*, 103(1), 21–38. <https://doi.org/10.1080/04353684.2021.1884497>
- Heo, C. Y., Kim, B., Park, K., & Back, R. M. (2022). A comparison of Best-Worst Scaling and Likert Scale methods on peer-to-peer accommodation attributes. *Journal of Business Research*, 148, 368–377. <https://doi.org/10.1016/j.jbusres.2022.04.064>
- Soft Skills Assessment Testing – Pymetrics*. (n.d.). Retrieved September 28, 2023, from <https://www.pymetrics.ai/>
- Jooss, S., Collings, D. G., McMackin, J., & Dickmann, M. (2023). A skills-matching perspective on talent management: Developing strategic agility. *Human Resource Management*. <https://doi.org/10.1002/hrm.22192>
- Kagan, C., & Burton, M. (2005). Marginalization. In G. Nelson & I. Prilleltensky, *Community Psychology* (pp. 293–308). Macmillan Education UK.  
[https://doi.org/10.1007/978-0-230-21400-2\\_14](https://doi.org/10.1007/978-0-230-21400-2_14)
- Kashdan, T. B., Disabato, D. J., Goodman, F. R., & McKnight, P. E. (2020). The Five-Dimensional Curiosity Scale Revised (5DCR): Briefer subscales while separating overt

and covert social curiosity. *Personality and Individual Differences*, 157, 109836.

<https://doi.org/10.1016/j.paid.2020.109836>

Lamri, J., & Lubart, T. (2023). Reconciling Hard Skills and Soft Skills in a Common Framework: The Generic Skills Component Approach. *Journal of Intelligence*, 11(6), 107. <https://doi.org/10.3390/jintelligence11060107>

Lavanchy, M., Reichert, P., Narayanan, J., & Savani, K. (2023). Applicants' Fairness Perceptions of Algorithm-Driven Hiring Procedures. *Journal of Business Ethics*. <https://doi.org/10.1007/s10551-022-05320-w>

Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L., Strong, D. R., & Brown, R. A. (2002). Evaluation of a behavioral measure of risk taking: The Balloon Analogue Risk Task (BART). *Journal of Experimental Psychology: Applied*, 8(2), 75–84. <https://doi.org/10.1037/1076-898X.8.2.75>

Leligou, H. C., Anastasopoulos, D., Vretos, N., Solachidis, V., Kantor, E., Plašilová, I., Girardet, E., Montagna, A., Vlahaki, F., & Tountopoulou, M. (2021). Experiences and Lessons Learnt from the Evaluation of ICT Tools for and with Migrants. *Social Sciences*, 10(9), Article 9. <https://doi.org/10.3390/socsci10090344>

Lewicki, K., Lee, M. S. A., Cobbe, J., & Singh, J. (2023). Out of Context: Investigating the Bias and Fairness Concerns of “Artificial Intelligence as a Service.” *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–17.

<https://doi.org/10.1145/3544548.3581463>

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 22 140, 55–55.



- Loughry, M. L., Ohland, M. W., & DeWayne Moore, D. (2007). Development of a Theory-Based Assessment of Team Member Effectiveness. *Educational and Psychological Measurement, 67*(3), 505-524. <https://doi.org/10.1177/0013164406292085>
- Microsoft for Startups (Director). (2017, January 30). *Pymetrics Internal Demo Day Pitch*. <https://www.youtube.com/watch?v=hzSImZZQZgQ>
- Nemoto, T., & Beglar, D. (2014). Developing Likert-scale questionnaires. In N. Sonda & A. Krause (Eds.), *JALT2013 Conference Proceedings*.
- Nielsen, J., & Molich, R. (1990). Heuristic evaluation of user interfaces. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 249–256*. <https://doi.org/10.1145/97243.97281>
- Paulsson, A., Runeson, P., & Ros, R. (2022). A/B Testing in the Small: An Empirical Exploration of Controlled Experimentation on Internal Tools. In D. Taibi, M. Kuhrmann, T. Mikkonen, J. Klünder, & P. Abrahamsson (Eds.), *Product-Focused Software Process Improvement* (pp. 449–463). Springer International Publishing. [https://doi.org/10.1007/978-3-031-21388-5\\_31](https://doi.org/10.1007/978-3-031-21388-5_31)
- Persson, H., Åhman, H., Yngling, A. A., & Gulliksen, J. (2015). Universal design, inclusive design, accessible design, design for all: Different concepts—one goal? On the concept of accessibility—historical, methodological and philosophical aspects. *Universal Access in the Information Society, 14*(4), 505–526. <https://doi.org/10.1007/s10209-014-0358-z>
- Predictive Hiring: How Artificial Intelligence is Helping Recruiters w/ @Kristen\_Hammy @JoinKoru (Episode 27) #DataTalk*. (2018). Experian Global News Blog. <https://www.experian.com/blogs/news/datatalk/predictive-hiring/>

- 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>
- Ramos-Villagrasa, P. J., Barrada, J. R., Fernández-del-Río, E., & Koopmans, L. (2019). Assessing Job Performance Using Brief Self-report Scales: The Case of the Individual Work Performance Questionnaire. *Journal of Work and Organizational Psychology*, 35(3), 195–205. <https://doi.org/10.5093/jwop2019a21>
- Rattray, J., & Jones, M. C. (2007). Essential elements of questionnaire design and development. *Journal of Clinical Nursing*, 16(2), 234–243.  
<https://doi.org/10.1111/j.1365-2702.2006.01573.x>
- Rentzsch, R., & Staneva, M. (2020). Skills-Matching and Skills Intelligence through curated and data-driven ontologies. *Proceedings of DELFI Workshops*.
- Robles, M. M. (2012). Executive Perceptions of the Top 10 Soft Skills Needed in Today's Workplace. *Business Communication Quarterly*, 75(4), 453–465.  
<https://doi.org/10.1177/1080569912460400>
- Rohrer, C. (2022). *When to Use Which User-Experience Research Methods*. Nielsen Norman Group. <https://www.nngroup.com/articles/which-ux-research-methods/>
- Romero-Díaz de la Guardia, J. J., García-Garnica, M., Chacón-Cuberos, R., & Expósito-López, J. (2022). Psychometric Validation of a Teamwork Skills Scale in a Vocational Training Context. *SAGE Open*, 12(2), 21582440221103256.  
<https://doi.org/10.1177/21582440221103256>

- Ruberg, B., & Ruelos, S. (2020). Data for queer lives: How LGBTQ gender and sexuality identities challenge norms of demographics. *Big Data & Society*, 7(1), 2053951720933286. <https://doi.org/10.1177/2053951720933286>
- Rubin, R. B., Palmgreen, P., Sypher, H. E., Rubin, R. B., Palmgreen, P., & Sypher, H. E. (2020). *Communicative Competence Scale* (R. B. Rubin, P. Palmgreen, & H. E. Sypher, Eds.; 1st ed., pp. 125–129). Routledge. <https://doi.org/10.4324/9781003064343-17>
- Sarigoz, O. (2012). Assessment of the High School Students' Critical Thinking Skills. *Procedia - Social and Behavioral Sciences*, 46, 5315–5319. <https://doi.org/10.1016/j.sbspro.2012.06.430>
- Sim, G., Read, J. C., & Holifield, P. (2008). *Heuristics for evaluating the usability of CAA applications*. Loughborough University. [https://repository.lboro.ac.uk/articles/conference\\_contribution/Heuristics\\_for\\_evaluating\\_the\\_usability\\_of\\_CAA\\_applications/9488765/1](https://repository.lboro.ac.uk/articles/conference_contribution/Heuristics_for_evaluating_the_usability_of_CAA_applications/9488765/1)
- Strachota, E. M., Conceição, S. C. O., & Schmidt, S. W. (2006). An Instrument Development Model for Online Surveys in Human Resource Development and Adult Education. *New Horizons in Adult Education and Human Resource Development*, 20(2), 24–37. <https://doi.org/10.1002/nha3.10249>
- Strom, P. S., & Strom, R. D. (2011). Teamwork skills assessment for cooperative learning. *Educational Research and Evaluation*, 17(4), 233–251. <https://doi.org/10.1080/13803611.2011.620345>
- Taylor, E. (1998). A Primer on Critical Race Theory. *The Journal of Blacks in Higher Education*, 19, 122. <https://doi.org/10.2307/2998940>

*The 7 Steps of the Engineering Design Process.* (2023). Indeed.

<https://www.indeed.com/career-advice/career-development/design-process>

Tountopoulou, M., Karaminas, N., Drosos, N., & Vlachaki, F. (2020). The NADINE Soft Skills Tests for Migrants, Refugees and Asylum Seekers: Cultural Adaptation and Preliminary Testing of Their Psychometric Properties. *Psychology*, 11(12), Article 12.

<https://doi.org/10.4236/psych.2020.1112121>

Tsirkas, K., Chytiri, A.-P., & Bouranta, N. (2020). The gap in soft skills perceptions: A dyadic analysis. *Education + Training*, 62(4), 357–377. [https://doi.org/10.1108/ET-03-](https://doi.org/10.1108/ET-03-2019-0060)

[2019-0060](https://doi.org/10.1108/ET-03-2019-0060)

Vasconcellos, Marcela. (2023). *Fair Many-to-Many Matching of Refugees to Jobs with Ties, Incomplete Lists, and Elastic Lower and Upper Quotas via Integer Optimization* [PowerPoint Slides].

*Welcome to CATME - Smarter Teamwork.* (2024). CATME. Retrieved September 27, 2023, from <https://info.catme.org/>

White, S. M., Riley, A., & Flom, P. (2013). Assessment of Time Management Skills (ATMS): A Practice-Based Outcome Questionnaire. *Occupational Therapy in Mental Health*. <https://www.tandfonline.com/doi/full/10.1080/0164212X.2013.819481>

## 9. Appendix

### Appendix A: Heuristic Protocol with Intrare Team Members

**Objective:** To assess the interface using the usability heuristics model with subject matter experts from the Intrare team.

**Participants:** Three to five Intrare team members who are subject matter experts.

**Procedure:**

- a. Provide subject matter experts with the usability heuristics model and the list of 13 heuristics which can be found in table 4.
- b. Ask participants to evaluate the skills assessment interface based on each of the 13 heuristics using the severity rating below.

**0** = I don't agree that this is a usability problem at all

**1** = Cosmetic problem only: need not be fixed unless extra time is available on project

**2** = Minor usability problem: fixing this should be given low priority

**3** = Major usability problem: important to fix, so should be given high priority

**4** = Usability catastrophe: imperative to fix this before product can be released

- c. Participants will report feedback with the provided document.
- d. After participants have evaluated the interface, facilitate a group discussion where participants discuss their ratings and reasoning.
- e. From the group discussion create a consolidated list of issues ranked by severity.

### Heuristics Evaluation Document

<b>Heuristics</b>	<b>Description</b>	<b>Severity Rating (0-4)</b>	<b>Comments</b>
1. There is clear language and grammar within questions	Text should be grammatically correct and make sense.		
2. Progress through the test is visible and understandable	Ensure that the number of questions answered and remaining is obvious and time remaining is clear.		
3. Answering questions is intuitive	Clear distinction between question styles and the process of answering the question should not be demanding.  Answering the question should be matched to interface components.		
4. Actions are easily reversed	It should be possible to change or remove an answer. Ensure it is possible to return to an incomplete test or question.		

<p>5. Users are informed of any unanswered questions before finishing</p>	<p>If a user has opted to end the test ensure that they are informed of any unanswered questions.</p>		
<p>6. Interface design characteristics are appropriate</p>	<p>Interface should match standards and design should support user tasks.</p>		
<p>7. Visual layout is appropriate - adequate spacing and visibility of questions</p>	<p>Ensure that there is enough spacing between the elements within the interface and scrolling is minimized within the questions.</p>		
<p>8. Feedback is appropriate</p>	<p>System feedback should be clear about what action is required.</p>		
<p>9. Moving between questions and terminating the exam is intuitive</p>	<p>User input to navigating between questions and returning to unanswered questions should be consistent. Options to exit should be identifiable.</p>		
<p>10. Time delays are minimized</p>	<p>Prevent any unnecessary delays. Ensure that there is minimal latency when moving between questions or saving answers.</p>		

<p>11. External influences to the user are minimized</p>	<p>Ensure test mode does not impact on fairness and performance within the test.</p> <p>Prevent distractions to other users and do not penalize them due to constraints of the software e.g. spelling mistakes (unless essential)</p>		
<p>12. The system matches the real world</p>	<p>Independent of the objective culture, the website should speak the language of its users with words, phrases and concepts familiar to them, making the possible actions easy to understand, instead of using concepts related to the system or technical terms. The website should use conventions of the objective culture such as: images, colors, and familiar objects, in a logical and natural order.</p>		
<p>13. Users can realize things in a consistent and standard way</p>	<p>Independent of the objective culture, the system should not make the users think if two actions, situations, or different words mean the same. Established conventions must be followed, under conditions that the user can realize</p>		



	things in a familiar, standard, and consistent way.		
--	---	--	--

## **Appendix B: A/B Testing Protocol**

**Objective:** To analyze the impact of different features of the prototype on the user experience and development of an engaging interface.

**Participants:** Any visitor of the Intrare website from December 11th, 2023 to January 8th, 2024 to optimize the number of participants.

### **Procedure:**

- a. The testing will run from December 11th, 2023 to January 8th, 2024. The experiment is based on this span of time with respect to the smaller use base in the hopes to maximize the number of responses.
  - i. The goal is to receive at least 60 responses (about 30 for each version), but ideally as many responses as possible.
- b. Users will be randomly assigned to each version where nearly half will be assigned to the A version (control) and the other half to the B (test).
- c. Users will complete the assessment and their interactions with the interface will be recorded.
- d. At the end of the study period, the results will be collected and analyzed.

# **Appendix C: Focus Group with Intrare Users Protocol**

**Objective:** Explore perceptions and attitudes towards the proposed skills assessment.

**Participants:** 5-8 current Intrare users

## **Procedure:**

### **1. Introduction:**

Hello everyone! Thank you for taking the time to join us today. The insight you provide will be beneficial and insightful to our team.

## **Purpose:**

We are here today to learn about your thoughts in relation to the Intrare application process. There is a new questionnaire which will help Intrare create a better match between users and companies. As we create this questionnaire, we want to make sure it reflects your own views and experiences.

## **Notes:**

Before we start, we want to encourage you to openly share your opinion and respect the privacy of others. Also, since this discussion is private, please do not repeat anything discussed here to your friends and family.

## **Agenda:**

Now we will review the agenda for today. First, we'll all introduce ourselves. Then we'll ask some questions about your thoughts on the questionnaires. Finally, we will open it up to everyone to hear any other feedback.

## **Thank you:**

We wanted to thank you again for taking the time to participate today. We look forward to the discussion that follows.

### **2. Discussion:**

**Introductions:**

- a. Have members of the group introduce themselves and say their favorite food (to serve as an icebreaker)
- b. What has your experience with Intrare been like in the past?
- c. Have you taken any online assessments in the past?
- d. Do you mainly use your phone to access the internet?
  - i. (Follow up, if needed) What are the other ways you access the internet?

**Interface:**

- a. Readability
  - i. What did you think about the font size of the text?
    1. (Follow up, if needed) Would you prefer if the font size was larger or smaller?
  - ii. How did you feel about the use of the color throughout the assessment?
    1. Did the color of the text impact your ability to read it?
    2. How did the use of color impact your ability to complete the assessment?
      - a. (Follow up, if needed) Did you find it more helpful or distracting to complete the assessment?
- b. Flow
  - i. What was most confusing about moving from question to question?
- c. Length/Progress
  - i. How did you feel about the length of the questionnaire?
  - ii. Did you feel encouraged throughout the process?
    1. Can you tell me more about that?
  - iii. Did you have a point where you stopped taking the assessment? If so, when?
    1. What made you want to quit?
    2. Is there anything that could be added which would make you want to finish the assessment?
- d. Overall clarity
  - i. How clear was the motivation behind the assessment?
    1. (Follow up, if needed) Were you able to understand how the assessment could help create job matches?

2. Did this make you want to take the assessment? Why or why not?
- ii. How did you feel about the questions being asked?
  1. Were any of the questions confusing or unclear?
- iii. Did the wording of the questions impact your motivation to finish the quiz?
- e. General feedback
  - i. How satisfied or dissatisfied are you with this process?
  - ii. Would you want to take this assessment again and why?
    1. (Follow up, if needed) Did you enjoy taking the assessment?
  - iii. What would you change if you had the opportunity?
    1. (Follow up, if needed) What elements did you like best and which did you like least?

**Questions:**

- a. Were any questions confusing or unclear?

**Feedback and Recommendations:**

- a. Open discussion

**3. Closing Question:**

In the last moments of this discussion is there anything else you would like to share or emphasize about the questionnaire?

**4. Closing Remarks:**

On behalf of our team, we wanted to thank everyone for participating in the discussion today. Your feedback will provide us with a strong understanding of what matters to users. Thank you again for taking time out of your day to participate in today's study.

## **Appendix D: App Translations**

### **Introduction: Spanish**

Por favor, complete este cuestionario sobre sus habilidades. Cuanto más honesto sea con sus respuestas, mejor se ajustará el trabajo a usted. Tomará aproximadamente 10 minutos, y puede tomarse el tiempo que necesite.

### **Introduction: English**

Please complete this questionnaire on your skills. The more honest you are with your answers, the better your job will fit you. It will take about 10 minutes, and you can take as long as you need.

### **Milestone 1**

1. I feel that my peers should respect my expertise.
  - a. Creo que mis compañeros deberían respetar mis especialidades.
2. I try to solve problems when I have them.
  - a. Intento solucionar problemas al encontrarlos.
3. I find it easy to communicate with others.
  - a. Tengo facilidad comunicándome con los demás.
4. I pay attention to what's happening around me.
  - a. Presto atención a lo que pasa a mi alrededor.
5. In group settings, I care about doing things well.
  - a. En equipos, me importa hacer las cosas bien.
6. I like having a daily routine.
  - a. Me gusta tener una rutina diaria.
7. I have a way to keep track of my schedule.
  - a. Tengo una manera de mantener mi horario.
8. I can solve problems and arrive at conclusions.
  - a. Puedo solucionar problemas y llegar a conclusiones.
9. I can handle situations, even if I don't have all the information.
  - a. Puedo manejar situaciones, aunque no tenga toda la información.
10. I care about what others think and say.

- a. Me importa lo que digan y piensen los demás.
- 11. I like to make sure I am on track to meet my goals with others.
  - a. Me gusta asegurarme que estoy en camino de alcanzar mis objetivos con los demás.
- 12. I like to plan and organize things with others.
  - a. Me gusta planear y organizar cosas con otra gente.
- 13. I can explain my thoughts logically and convincingly.
  - a. Puedo explicar mis pensamientos de manera lógica y convincente.
- 14. I see tough situations as chances to learn and grow.
  - a. Considero situaciones difíciles como oportunidades para aprender y crecer.
- 15. I can change how I talk based on the situation.
  - a. Puedo cambiar mi forma de hablar dependiendo de la situación.
- 16. I meet the goals that I set for myself.
  - a. Logro las metas que me propongo.
- 17. I make sure to finish my tasks on time.
  - a. Me aseguro de acabar mis tareas a tiempo.
- 18. My job is part of my identity.
  - a. Mi trabajo es parte de mi identidad.

## **Milestone 2**

- 19. I give myself enough time to finish my tasks.
  - a. Me doy tiempo para completar mis tareas.
- 20. In group settings, I try to do my fair share of work.
  - a. En grupos, intento trabajar de manera justa.
- 21. I enjoy figuring out how things work.
  - a. Me gusta averiguar como funcionan las cosas.
- 22. I feel capable of dealing with everyday challenges.
  - a. Soy capaz de superar dificultades del día a día.
- 23. I celebrate when my teams do well.
  - a. Celebro cuando mis equipos lo hacen bien.
- 24. I am able to set my priorities.
  - a. Soy capaz de tener prioridades.

25. I'm good at managing my time.  
a. Soy bueno gestionando mi tiempo.
26. In discussions, I listen to everyone and respect their opinions.  
a. En las discusiones, escucho lo que todos tienen que decir y respeto sus opiniones.
27. I am relaxed and comfortable when talking in front of others.  
a. Soy alguien relajado y no tengo dificultad para hablar en público.
28. I take on extra tasks and challenges when I can.  
a. Me encargo de tareas extras cuando puedo.
29. I ask people for their opinions before making decisions.  
a. Antes de decidir algo, hablo con gente para saber lo que opinan.
30. I do things based on my feelings and others' feelings.  
a. Tomo decisiones dependiendo de mis sentimientos y los de los demás.

### **Milestone 3**

31. I pay attention when I talk to people.  
a. Presto atención al hablar con alguien.
32. I pay attention when others talk.  
a. Presto atención y escucho cuando alguien habla.
33. If I saw something was done incorrectly I would report it to the proper authority.  
a. Si veo que algo se está haciendo mal, lo denunciaría a la autoridad correspondiente.
34. I like to solve complex problems.  
a. Me gusta solucionar problemas complicados.
35. I'm always on time when meeting others.  
a. Siempre soy puntual al quedar con gente.
36. I manage my emotions to avoid making situations worse.  
a. Tengo control de mis emociones para evitar que las situaciones empeoren.
37. I find it interesting to learn new things.  
a. Me parece interesante aprender cosas nuevas.
38. I tell my peers if they make mistakes.  
a. Si alguien comete algún error, no me cuesta decírselo.
39. I enjoy learning about new topics.  
a. Me gusta aprender cosas nuevas.



40. I keep an open mind when listening to what others have to say.

a. Mantengo mi mentalidad abierta al hablar con alguien.

41. I feel confident about making a positive impact in any setting.

a. Estoy seguro que soy capaz de tener un efecto positivo en cualquier situación.

42. I look for hope even in tough times.

a. Tengo esperanza hasta en situaciones difíciles.

**Conclusion: Spanish**

¡Buen Trabajo! Usted acaba de terminar el cuestionario. Por favor, haga clic en el botón de abajo para ver sus resultados.

**Conclusion: English**

Good job! You finished the questionnaire. Please click on the button below to view your results.

## Appendix E: Encouragements

The following is a list of encouragements as well as an indication of which question they are displayed on.

Question: 4

Message: "¡Buen Trabajo!"

Translation: "Good job!"

Question: 9,

Message: "¡Sigue así!"

Translation: "Keep going!"

Question: 13,

Message: "¡No te rindas!"

Translation: "Don't give up!"

Question: 19,

Message: "¡Terminaste la etapa 1!"

Translation: "You finished stage 1!"

Question: 22,

Message: "¡Estás a mitad de camino!",

Translation: "You're halfway there!"

Question: 25,

Message: "¡Vas muy bien!",

Translation: "You're doing really well!"

Question: 31,

Message: "¡Terminaste la etapa 2!",

Translation: "You finished stage 2!"

Question: 37,

Message: "¡Sí se puede!"

Translation: "You can do it!"

Question: 39,

Message: "¡Estás tan cerca!"

Translation: "You're so close!"