Journal of Economics, Business & Organization Research

Year: 2023, Vol: 5, Issue: 2, pp. 110-133

Article Type: *Research*

Citation: Cavus, B., Kazancı, O., & Esen, M. (2023). Investigating the reliability of ChatGPT in assessing job satisfaction. *Journal of Economics, Business and Organization Research, 5*(2), 110-133.

INVESTIGATING THE RELIABILITY OF CHATGPT IN ASSESSING JOB SATISFACTION

Burak CAVUS[1]
Onur KAZANCI[2]
Murat ESEN[3]

Received: 29.05.2023 **Accepted:** 17.08.2023 **Published Online:** 29.10.2023

Abstract

Today, there is increasing controversy surrounding the use of artificial intelligence software and programs. Although many organizations are attempting to boost their performance by utilizing the potential of artificial intelligence, controversy about the technology's reliability prevails. This study investigates the reliability of ChatGPT in measuring job satisfaction. The Minnesota Job Satisfaction Questionnaire (MSQ), a reliable scale, and ChatGPT results were compared using the Pearson correlation coefficient Correlation technique. The SPSS 25 analysis findings revealed a strong connection between the two questionnaires. This proves that ChatGPT is a reliable tool for assessing job satisfaction. The ability of the questionnaire to correctly predict the MSQ results provided more evidence of the questionnaire validity. However, the study was limited in its sampling in terms of variety. Additionally, the sample size of the study was relatively small and restricted to almost one organization. The results of this study point to ChatGPT as a potential technology for gaging employee engagement and job satisfaction in businesses. These findings may have consequences for further research on these topics.

Keywords: *ChatGPT, Job Satisfaction, MSQ, Organizational Behaviour, Artificial Intelligence*

[1] PhD Candidate, Izmir Katip Celebi University, Institute of Social Sciences, Department of Business Administration, Türkiye, burak.cavus@ikcu.edu.tr, Orcid: 0000-0001-8112-8134
[2] Asst. Prof., İzmir Katip Çelebi University, Faculty of Economics and Administrative Sciences, Department of Business Administration, Türkiye, onur.kazanci@ikcu.edu.tr, Orcid: 0000-0001-6242-2637
[3] Assoc. Prof., Kyrgyz-Turkish Manas University, Faculty of Economics and Administrative Sciences, Department of Business Administration, Kyrgyzstan, murat.esen@manas.edu.kg, Orcid: 0000-0003-2434-6586

1. INTRODUCTION

Since the inception of work itself, there has been a notion of job satisfaction. Because of its importance in influencing employee motivation and productivity, job satisfaction has been extensively researched. The concept of job satisfaction and motivation in the workplace was first mentioned by pioneers in the late 19th century, including Weber (1947), Taylor (1911), and Mayo (1933). Since then, many academics have worked to determine the main elements that affect job satisfaction and to clarify the connection between job performance and job satisfaction.

Depending on the individual and the particular job, the significance of these dimensions can vary, but they are widely regarded as some of the major variables that can impact job satisfaction.

In recent years, a growing number of people have been interested in using artificial intelligence (AI) to evaluate job satisfaction. Our study topic focused on whether ChatGPT might be used to quantify job satisfaction through natural language processing. By contrasting its findings with those of the Minnesota Job Satisfaction Questionnaire (MSQ), this study will assess the validity of the ChatGPT questionnaire. The correlation between the two surveys will be examined using the Pearson correlation method, and the data will then be analysed using SPSS 25. This will enable us to evaluate the validity of the ChatGPT survey in gaging job satisfaction.

2. LITERATURE REVIEW

2.1. Job Satisfaction

The field of organizational behaviour and economic theory employs the concept of job satisfaction to elucidate the intricate correlation between employee well-being and production. This phenomenon encompasses a multifaceted combination of affective, cognitive, and behavioural elements that reflect an employee's level of contentment with their job and associated duties. For example, in classical economic theory, individuals strive to optimize their utility, encompassing both material prosperity and personal contentment. Moreover, within the framework of institutional economics, the presence of regulations and adherence to social norms create a conducive atmosphere that ensures equitable treatment of workers, their active involvement in decision-making processes, and safeguards them against any form of exploitation. Furthermore, behavioural economics acknowledges that individuals' decision-making processes are subject to the influence of psychological factors. In most of the theories, job satisfaction finds a vital seat for organizational efficiency.

The level of job satisfaction has a significant impact on employee turnover, absence rates, and the development of skills. The concept of job satisfaction is analysed through economic frameworks such as utility maximization and hedonic well-being, which contribute to our comprehension of individuals' preferences and behaviours in the labour market.

In the field of organizational behaviour research, job satisfaction is a crucial subject. It is commonly known that employee performance, productivity, and morale impact job satisfaction directly (Harter et al., 2002). The elements that affect job satisfaction have been examined by numerous academics, including job autonomy (Hackman and Oldham, 1975a: 165), job security, and job role (Ostroff, 1992: 965). Additionally, research has shown that factors like age, gender, and occupation have an impact on job satisfaction (Harter et al., 2002).

Research has indicated that organizational factors such as pay and benefits (Aamodt, 2014: 337), the work environment (Hackman and Oldham, 1976b: 255), and job security (Ostroff, 1992: 970) affect job satisfaction in addition to personal characteristics. Additionally, it was hypothesized that psychological factors such as job-related stress and job burnout (Demerouti et al., 2005: 139) could have an impact on job satisfaction. Finally, some research has shown that organizational culture, including teamwork and communication, has an impact on job satisfaction (Chatman and Barsade, 1995: 335).

According to Locke (1976), "a pleasurable or positive emotional state resulting from the appraisal of one's job or job experiences" is what is meant by job satisfaction. Job satisfaction, in Locke's view, is a result of the gap between what an individual expects from a job and what they believe they are receiving in return.

The Hawthorne Studies were studied by Elton Mayo and his colleagues in the 1900s, are among the most frequently recognized studies on job satisfaction (Mayo, 1933). According to this study, worker productivity increased because they were being monitored and paid attention, not because of changes in working circumstances. The human relations approach to management, which placed a strong emphasis on treating employees more like people than just tools in a machine, was developed as a result of this study.

The bidirectional causality between job satisfaction and productivity has been empirically demonstrated employing rigorous econometric approaches (Oswald et al., 2015: 807).

Therefore, the integration of job satisfaction into economic theory enriches the examination of the labour market and underscores its significance in cultivating a robust and efficient workforce.

In their study, Hakanen and Schaufeli (2012) investigated the link between burnout, work engagement, depressive symptoms, and life satisfaction over a span of seven years. According to this study, there is a positive correlation between increased work engagement and decreased burnout, which in turn relates to a reduction in depressive symptoms and an increase in overall life satisfaction (Hakanen and Schaufeli, 2012: 422).

A recent study in higher education institutions in Vietnam investigated the correlation between internal communication, employee engagement, job satisfaction, and employee loyalty. The enhancement of internal communication has been found to positively impact employee engagement, job satisfaction, and loyalty (Nguyen et al., 2023: 11).

Below are a few prominent recent studies on job satisfaction;

1. Income comparisons can influence job satisfaction and well-being (Clark and Senik, 2016).
2. Perceptions of organizational politics impact job satisfaction and job performance (Rosen et al., 2016).
3. The Job Demands-Resources (JD-R) theory links job characteristics, resources, and job satisfaction (Bakker and Demerouti, 2017).
4. Saks (2017) explored the relationship between employee engagement, job satisfaction, and performance outcomes (Saks, 2017).
5. The role of psychological capital (PsyCap) in enhancing employee well-being and performance (Luthans and Youssef-Morgan, 2017).
6. Employees' proactive adjustments to their job tasks (job crafting) influence job satisfaction and well-being (Tims et al., 2019).
7. The mediating role of work engagement in the relationship between job characteristics and employee behaviours (Sulea et al., 2015).

Based on the collective findings of several research, factors such as pay, political decisions by administration, job characteristics and resources, engagement, and performance outcomes, PsyCap, and job crafting have a direct or indirect impact on employees' job satisfaction. Expanding upon these observations, the measurement of job satisfaction emerges as a pivotal undertaking in comprehending the complex interaction of different variables within the organizational setting.

To adequately assess job satisfaction, it is essential to adopt an approach. In this regard, quantitative methods encompass the use of standardized surveys to evaluate numerous components of job satisfaction, including salaries, benefits, working conditions, opportunities for advancement, and interpersonal dynamics.

2.2. Most Used Job Satisfaction Scales

To our knowledge, the main job satisfaction scales commonly used in research and practice are listed below;

Job Satisfaction Questionnaire (JSS): The JSS is developed by Paul Spector and is a very commonly used scale to assess job satisfaction across multiple facets, including pay, promotion opportunities, supervision, co-workers, and the work itself. It consists of 36 items and has been validated in various work settings and populations (Spector, 1985). JSS has demonstrated good reliability and validity in numerous studies and has been used in various research on job satisfaction.

Minnesota Satisfaction Questionnaire (MSQ): The MSQ, developed by David J. Weiss, is another popular scale that assesses job satisfaction across dimensions such as work itself, supervision, colleagues, pay, and chances for advancement. It has two versions - the long form with 100 items and a shorter version called the MSQ-Short Form (MSQ-SF) with 20 items (Weiss et al., 1967). The MSQ has been widely used in research and has good psychometric properties in various populations.

Job Descriptive Index (JDI): Smith, Kendall, and Hulin developed the JDI. It is a widely used scale that measures job satisfaction based on five facets: work itself, pay, promotion chances, supervision, and colleagues. It consists of 72 items and is frequently used in research and organizational contexts (Smith et al., 1969). Because it is valid and reliable, JDI has been used in numerous studies to gage job satisfaction.

Scale for Jobs in General (JIG): The JIG Scale is a global indicator of job satisfaction that measures overall job satisfaction. Ironson, Smith, Brannick, Gibson, and Paul developed it (Ironson et al., 1989). Participants must rate their degree of job satisfaction on a scale of 1 to 7. The higher scores prove a better level of job satisfaction. It is quick and simple to determine job satisfaction using the JIG Scale. It is routinely incorporated into longer surveys and has been used in various studies.

The Faces Scale: The faces scale is a visual scale that shows respondents a series of faces that, from very joyful to very unhappy, represent various levels of job satisfaction. Selecting the face that most accurately reflects a respondent's degree of job satisfaction is required. This scale is straightforward and simple to use, and it is frequently used in studies or questionnaires with few possibilities for responses (Kunin, 1955).

Brief Index of Affective Job Satisfaction (BIAJS): The Weiss, Dawis, England, and Lofquist team created the Brief Index of Affective Job Satisfaction (BIAJS), a short and straightforward scale that gages affective (emotional) job satisfaction based on six items that evaluate positive and negative effect related to a person's job (Weiss et al., 1967). It has been proved to have valid psychometric qualities and has been used in many studies to evaluate emotional responses to one's job.

Researchers should carefully choose the scale that best suits their needs based on their specific aims, demographics, and research design in order to consider the possibility that alternative scales may be more appropriate for various research or organizational situations. In-depth descriptions of these scales' psychometric characteristics and further information are available in the original references listed. In this study, we chose MSQ due to the volume of questions being appropriate for study.

2.3. The development of Artificial Intelligence

Artificial intelligence (AI) refers to a machine's capacity to perform operations that are ordinarily specific to human intelligence, such as understanding natural language, recognizing and identifying objects, and making judgment calls. The earliest attempts to create robots that could think and reason as if they were real people were achieved in the 1950s, which is when the history of AI really began to take shape.

With their seminal paper "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence" (McCarthy et al., 1955), pioneers like John McCarthy and Marvin Minsky laid the groundwork for the field in the first development of AI studies.

The years that followed saw the development of numerous AI methodologies, such as symbolic logic, rule-based systems, and neural networks. One of the most significant early AI systems was the General Problem Solver (GPS), developed by Allen Newell and Herbert Simon (Newell and Simon, 1961). This system showed how computers might handle various problems by following a set of rules.

Despite these early developments, the initial creation of AI systems did not perform well enough, and funding problems caused a slowdown in AI research in the 1970s. This phase is known as the "AI winter," and marked uncertainty and a loss of belief in the sector.

However, AI technologies started to emerge again in the 1990s because of developments in machine learning and the availability of more powerful computers. One of the most important key developments was the development of artificial neural networks, which allowed computers to learn from data and improve their performance. The backpropagation algorithm developed by Rumelhart, Hinton, and Williams in 1986 had a major influence on the development of neural networks and the emergence of the deep learning concept (Rumelhart et al., 1986).

Another crucial development in the 1990s was the development of statistical natural language processing (NLP). This development allowed computers to understand and produce a human language. This was proved by the invention of the Hidden Markov Model for speech recognition and the probabilistic context-free grammar for natural language parsing (Rabiner, 1989; Pereira and Schabes, 1992).

Due to advancements in machine learning, computer vision, robotics, and natural language processing, research and development in artificial intelligence (AI) have accelerated significantly in the twenty-first century. One of the most significant recent breakthroughs has been the development of deep learning, which has led to improvements in speech recognition, image identification, and natural language understanding.

One of the most well-known deep learning models, the Generative Pre-trained Transformer (GPT) developed by OpenAI, has been used for many NLP tasks, including language translation, question answering, and text generation (OpenAI, 2023a). The most recent version of this model, GPT-4, which is currently in development, is reported to be significantly more powerful and versatile than its predecessors in terms of the number of data it is predicted to process. According to reports, GPT-4 can store 45 gigabytes of training data as opposed to 17 gigabytes of gigabytes.

Below is the timeline for developing AI technologies until GPT-4;

Early AI research started in the 1950s, leading to the creation of tools such as the Logic Theorist and the General Problem Solver. John McCarthy named the phrase "artificial intelligence" in 1956 at the Dartmouth Conference. In the 1960s, the ELISA chabot and the SHRDLU natural language processing program were two examples of rule-based AI systems that were created. Between 1970s–1980s, an era dubbed the "AI winter" resulted from a slowdown in AI research progress. Afterwards, expert systems and machine learning methods such as neural networks and decision trees were developed in the 1980s and 1990s. The development of probabilistic context-free grammar and the Hidden Markov Model are examples of statistical natural language processing approaches that were seen until 2000, and the idea of big data and developments in machine learning in the 2000s paved the way for innovations in computer vision and speech recognition. During 2010s, deep learning techniques such as convolutional neural networks and recurrent neural networks have transformed AI research and led to ground-breaking improvements in image and natural language processing. As we approach the 2020s, AI is still evolving and progressing with new developments in fields including robotics, reinforcement learning, and explainable AI. The most recent advancements were the processing of enormous amounts of data as a language model in GPT-3 and GPT-4. AI can disrupt numerous industries, affect how people live and work, as well as how organizations and businesses run as it continues to develop and advance.

2.4. AI Technologies in Use Today

In the contemporary era characterized by swift advancements in technology, AI has emerged as a catalyst for profound change, fundamentally altering various industries and the daily encounters of individuals. AI technology have been seamlessly incorporated into different facets of our daily lives, ranging from personalized recommendations on streaming platforms to the navigation of autonomous vehicles on our roadways. By utilizing sophisticated algorithms, machine learning techniques, and data processing skills, it empowers organizations to enhance their operational efficiency, enables academics to delve into intricate challenges, and allows individuals to engage with technologies in manners that were previously futuristic. Below we examined many prominent artificial intelligence (AI) technologies presently employeed, demonstrating their wide-ranging uses and significant influence on moulding both our current circumstances and future prospects.

Natural Language Processing (NLP): This is a technology that uses machine learning methods to help computers understand and interpret the human language. It is used in chariots and voice assistants among other applications.

Computer Vision: This technology, which is being used in applications such as facial recognition and self-driving cars, uses algorithms to provide computers with the ability to comprehend and evaluate visual data from photographs and videos.

Robotics: This technology uses intelligent machines that can complete activities on their own or with little assistance from a human being. It is employeed in industries including manufacturing and healthcare.

Expert Systems: These are computer programs that apply artificial intelligence methods to address issues in certain fields, including financial analysis or medical diagnostics.

Recommender Systems: These are algorithms used in applications such as e-commerce and streaming services that use information about a user's previous activity to propose goods or services they might be interested in.

Machine Learning: This area of artificial intelligence uses algorithms to teach computers how to learn from data and improve over time.

Deep Learning: Artificial neural networks are used in this area of machine learning to help computers learn from massive amounts of data and make predictions or judgments based on that data.

Reinforcement Learning: In this field of machine learning, algorithms that learn via trial and error are employeed, among other things, in robots and video games.

Generative Models: These algorithms are used in applications such as text and image production because they may create new data similar to current data.

These are only a handful of the numerous AI technologies that are now being created and employeed, and as new methods and uses are discovered, the area of AI continues to expand quickly.

2.5. Use of AI Technologies in Business

The strategic incorporation of AI technologies has emerged as a crucial tool for innovation and productivity in the contemporary business environment. It has significantly transformed organizational operations, decision-making processes, and stakeholder engagement. This revolution encompasses various aspects, such as the optimization of supply chain operations via the utilization of predictive analytics and the enhancement of consumer experiences through the implementation of chatbots and recommendation systems. Through the utilization of machine learning, natural language processing, and automation, enterprises have the ability to extract valuable insights from extensive datasets, optimize workflows, and attain a competitive advantage within a progressively data-centric global landscape.

Finance: AI technologies can be utilized in the financial sector to automate procedures, spot fraud and offer investment advice. For instance, Bridgewater Associates employs AI algorithms to decide which investments to make based on a large amount of data.

Retail: By offering products, customizing marketing messaging, and forecasting demand, AI technology in retail can enhance the customer experience. For instance, Amazon uses machine learning to offer products to users based on their browsing and purchase histories.

Production: AI technologies can be utilized to enhance quality control, monitor equipment performance, and optimize production processes. To forecast maintenance requirements and minimize downtime, General Electric, for instance, uses AI algorithms to evaluate data from sensors in its aircraft engines.

Healthcare: By evaluating medical data to diagnose diseases, identify risk factors, and offer treatment choices, healthcare AI technology can be used to enhance patient outcomes. As an example,IBM Watson used machine learning for Oncology to analyse medical data and offer individualised cancer treatments.

Transportation: By evaluating traffic patterns, streamlining routes, and anticipating maintenance requirements, transportation AI technology can be used to increase safety and efficiency in the industry. Uber, for example, uses machine learning algorithms to forecast demand and enhance driver routes.

Marketing: By evaluating customer data to tailor messaging, predict customer behavior, and optimize campaigns, marketing AI technology can be used to increase the effectiveness of marketing activities. For instance, Persado, a marketing platform, employs AI to create marketing messages that are tailored for particular consumers.

These are but a few instances of how AI technologies might be applied in various industries. As the technology develops, so do the potential applications of AI, which are numerous and only set to increase.

2.6. ChatGPT (GPT-3)

A cutting-edge new technology called ChatGPT has the power to transform how companies communicate with their clients and staff. Generalized Pre-trained Transformer (GPT3) is the name of the platform on which this technology was developed by a business called OpenAI. It is an innovative artificial intelligence (AI) system that can produce conversations that sound like human speech from a small quantity of data. The system can produce insightful responses to customer inquiries because it has been trained on various discussions.

It is frequently used in natural language processing (NLP) tasks such as question-answering, language synthesis, and summarization. GPT-3 is a potent language model. It boasts amazing powers in language recognition and generation because it is the largest language model ever created and has been trained on an enormous quantity of data (Dale, 2021). It can function well in a range of languages thanks to its demonstrated outstanding multilingual capacities (Armengol-Estap'e et al., 2021). This gives GPT-3 significant advantage because it makes wide variety of applications possible. Additionally, it has been demonstrated that GPT-3 is capable of producing text that is both engaging and natural (Miotto et al.,, 2022).

ChatGPT has been around for a couple of years, but only recently has it started to gain popularity. This system can produce conversations with sincere clients and human Since then, several companies have been using ChatGPT to automate customer communications, customer support processes, and internal processes such as employee satisfaction.

ChatGPT has several possible commercial uses (Hernandez, 2020). It can be used to automate customer service tasks such as answering inquiries, resolving issues, and providing product information. It can also be used to automate customer surveys and improve the client experience. Businesses can make customized recommendations to customers using ChatGPT based on their past behavior. Businesses could strengthen customer loyalty and increase revenue by doing this.

Two new products from the company were made available today: ChatGPT Plus, which provides internet access, and GPT-4, which is a more successful version of GPT-3.

2.7. GPT 4

On March 14th, 2023, OpenAI launched its newest version of GPT, known as GPT-4, which is claimed to be a better version of GPT-3. However, for the time being, there is still a waitlist and it is unavailable for everyone to test it. According to OpenAI, GPT-4 has made significant improvements in three areas compared to GPT-3.5. First, GPT-4 can read images and can provide logical answers to questions based on the content of the image. Secondly, GPT-4 is more accurate than its predecessor due to its larger transformer architecture, more complex neural network, and bigger training data set, which enable it to handle various tasks with better information processing. For example, in a mock American Bar Examination, GPT-4 outperformed GPT-3.5 by a significant margin. Lastly, GPT-4 supports longer inputs, which allows it to accept up to 32,768 tokens or about 25,000 words at a time. This helps it generate more coherent and natural text (Cheng et al., 2023).

OpenAI announced an extension of their partnership with Microsoft, which will allow them to continue their independent research and development of more safe, useful, and powerful AI (OpenAI, 2023b). According to the OpenAI website, a multi-billion dollar investment from Microsoft was made into OpenAI's research and development of AI. OpenAI is a capped-profit company that is governed by a non-profit organization to ensure that its core beliefs about sharing benefits and prioritizing safety are not sacrificed while raising the capital necessary to fulfill its mission. Microsoft's investment will help OpenAI continue its independent research and develop AI that is increasingly safe, useful and powerful. Microsoft has been instrumental in OpenAI's progress, providing supercomputing systems powered by Azure that have been crucial in delivering best-in-class performance and scale for AI training and inference workloads. OpenAI and Microsoft collaborate to review and synthesize shared lessons to build and deploy safe AI systems, and they have deployed OpenAI technology through their API and the Azure OpenAI Service (OpenAI, 2023a).

3. PURPOSE OF THE STUDY

The aim of this study was to assess the reliability of ChatGPT in measuring job satisfaction. Job satisfaction is an important factor in maintaining the performance and productivity of an employee. Therefore, it is essential to have reliable measurement tools to evaluate job satisfaction. We used parallel forms reliability test with the Pearson correlation coefficient method in SPSS to measure the correlation between the ChatGPT questionnaire and the Minnesota Job Satisfaction Questionnaire (MSQ). We also checked internal consistency reliability with Cronbach’s Alpha score. Cronbach's Alpha is a method used to test the internal consistency reliability of a test or scale. It is based on the average correlation among all possible pairs of items within the test. Cronbach's Alpha measures the degree to which items in the test are correlated with each other and provides a reliable measure of the underlying construct being measured. It ranges from 0 to 1, with higher values indicating greater internal consistency reliability (Streiner, 2003).

Cronbach's Alpha is a more general and sophisticated method for estimating internal consistency reliability than split-half reliability. Unlike split-half reliability, which divides the test into two halves, Cronbach's Alpha can handle tests with items that are not necessarily equivalent or homogeneous. Therefore, Cronbach's Alpha is a widely used method to evaluate the internal consistency of psychological tests and scales (Streiner, 2003).

To calculate Cronbach's Alpha, the scores of individuals on all items in the test are first summed to obtain a total score. Then, the correlation between each item and the total score is calculated. Finally, the average of all these correlations is computed to obtain the Alpha coefficient.

4. METHOD

4.1. Research Model

It was tried to test the reliability of ChatGPT’s job satisfaction questionnaire with the help of internal consistency (Nunnally and Bernstein, 1994) and parallel form reliability tests (Hilger and Beauducel, 2017) and the Pearson coefficient correlation method (Judge et al., 2001). This model is suitable for examining the correlation between the questionnaire prepared by the ChatGPT and the Minnesota Job Satisfaction Questionnaire (MSQ). It allows us to measure the strength and direction of the relationship between the two questionnaires and the ChatGPT questionnaire within itself.

4.2. Participants

The convenience sampling method was used in this study because it is a cost-effective and time-efficient way to collect data. Furthermore, convenience sampling is well suited for quick studies that have limited resources and time (Babbie, 2016). In this study, convenience sampling allowed us to quickly and easily collect the data that we needed to assess the reliability of ChatGPT in assessing job satisfaction. The sample size was determined according to the study objectives and requirements. 91 participants joined the research. 75 of them from the same organization and 16 from other organizations. All participants were asked to solve MSQ first and then the ChatGPT questionnaire. Detailed information about demographic variables related to work experience is given in Table1.

Table 1. Demographics (Work experience in years)

Years	1 – 5	6 – 10	11 – 20	More than 20
N of Participants	23	11	53	4

4.3. Data Collection Tool

The data collection tool used in this research was the Minnesota Job Satisfaction Questionnaire (MSQ) and the questionnaire prepared by ChatGPT. These two questionnaires were designed to measure job satisfaction among participants. The MSQ is a widely used questionnaire tool to assess job satisfaction. Developed in the 1970s by researchers at the University of Minnesota, the MSQ comprises 20 items that measure job satisfaction along dimensions such as education and career, compensation and reward, location and commute, workload and work convenience, work environment, and work climate. The questionnaire also includes an overall job satisfaction score, which is used to assess the overall satisfaction of an employee with their job. Organizations use the MSQ to assess employee job satisfaction and identify areas of concern that can be addressed to improve employee morale and productivity. In both questionnaires, the Likert scale is used out of five options.

To have a valid questionnaire similar to MSQ, a standard procedure was applied. In each step, results were sent to 2 expert lecturers in the field of organizational behaviour, and they analysed the reliability of the questions to measure job satisfaction. For the first attempt, ChatGPT was asked to create a set of questions to measure job satisfaction. The result questions were sent to expert lecturers but not found valid enough to measure job satisfaction. For the second attempt, ChatGPT was asked to create a set of questions concerning job satisfaction dimensions such as education and career, compensation and reward, location and commute, workload and work convenience, work environment, and work climate. The results were sent to expert lecturers. However, most of the questions were not found to be reliable and related to our purpose. Then, we tried to create a set of questions that expert lecturers elected for the latest questionnaire. For the last attempt, we extended our prompt to ChatGPT, and the following statement was prompted into the system; *"I am a researcher trying to measure job satisfaction in an institution. I would like to ask questions about dimensions such as education and career, compensation and reward, location and commute, workload and work convenience, work environment, and work climate. Prepare a set of questions to determine their effect on job satisfaction."*

The result questions were taken from ChatGPT and sent to expert lecturers. They applied various elections for these questions. Repeated questions, non-related questions, and questions that are not valid enough to measure job satisfaction were omitted and 20 questions were selected for our experiment (ChatGPT, 2020).

4.4. Data Analysis

The data were then analysed using SPSS 25 to measure the Pearson correlation coefficient. To apply the parallel form reliability test, we had the same number of questions in both questionnaires. These questions were analysed with a Varimax Rotated Component Matrix to determine the main dimensions in the questionnaire created by ChatGPT. These dimensions are then correlated with the human-made questionnaire MSQ.

4.5. Questionnaire Statistics

Table 2. Descriptive Statistics

Mean	Variance	Std. Deviation	N of items
69,80	209,072	14,459	20

According to the results from SPSS22, the mean was 69.80. This refers to the average score of the scale. In this study, the mean score is 69.80, indicating that the respondents' scores are centred around this value.

Our variance is 209.072. This refers to the measure of how spread out the data is around the mean. The data points are more dispersed as indicated by a bigger variance, while a lower variance indicates that the data points are closer together. In this study, the variance is 209.072, indicating that there is some variability in the scores of the respondents. This proves that some respondents may have scored significantly higher or lower than the mean.

Our survey standard deviation is 14.459. This refers to the amount of variability or dispersion of the data around the mean. When the standard deviation is low, the data points are concentrated around the mean, whereas when the standard deviation is high, the data points are dispersed more widely. The standard deviation in this instance was 14.459, indicating some variations in the respondents' scores.

We had 20 questions in the ChatGPT survey indicating that the scale covers a broad range of topics or concepts.

Overall, the results proves that the respondents' scores on the scale exhibit some variability, indicating that there may be differences in their responses to the items in the scale. However, the mean score of 69.80 proves that on average, the respondents scored close to this value. The standard deviation indicates that there is some variability in the scores, but not an extreme amount. The number of questions in the survey proves that the scale covers a broad range of topics or concepts.

4.5. Validity and Reliability

The results from SPSS 25 showed that there was a considerable correlation between two questionnaires, which indicates that ChatGPT can be reliable. The correlation was significant at almost all categories, which can be seen in the findings. The internal consistency reliability score from SPSS result can be seen in Table 3. It can be seen that the ChatGPT questionnaire had a positive internal consistency reliability score, which means that the questions in it can be regarded as internally correlated and meaningful for the participants.

Table 3. Reliability Statistics (Cronbach’s Alpha)

Cronbach's Alpha	N of Items
,940	20

Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy: 0.831. The KMO test measures the adequacy of the sample size for factor analysis. It ranges from 0 to 1 with values closer to 1 indicating a better sample size for factor analysis (Field, 2018). In this case, the KMO measure of 0.831 indicates that the sample size is considered to be "meritorious" for factor analysis, proving that there is a sufficient amount of correlation between the variables.

Bartlett's Test of Sphericity assesses whether the correlation matrix of the variables is significantly different from the identity matrix. A significant result ($p < 0.05$) proves that the data are appropriate for factor analysis (Field, 2018). In this case, the approximate chi-square value is 1580.124 with 190 degrees of freedom and a p-value of 0.000, indicating that the correlation matrix of the variables is significantly different from the identity matrix. Therefore, the result is considered significant, proving that the data are appropriate for factor analysis.

Overall, the results shown in Table 4 prove that the survey has acceptable construct validity, and that the data are appropriate for factor analysis. The KMO value of 0.831 indicates that there is sufficient correlation between the variables, while the significant Bartlett's test result indicates that the correlation matrix of the variables is significantly different from the identity matrix. These results provide some confidence that the survey has sufficient construct validity and can be used to examine underlying factors or dimensions.

Table 4. Construct the Validity of ChatGPT Questionnaire

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,831
Bartlett Test of Sphericity	**Approx. Chi-Square**	1580,124
	df	190
	Sig.	,000

4.6. Dimension of the questions

According to Varimax Rotated Component Matrix data driven from SPSS, we detected 4 dimensions in the questionnaire. These dimensions can be seen in Tables 5 and 6. It should be noted here that MSQ has 20 dimensions in its wider form and the dimensions listed here can be found within MSQ dimensions.

Table 5. Main Dimensions and Questions

Questions	Dimension
1. Are you generally satisfied with the location of your work?	General Satisfaction
6. How satisfied are you with the overall job?	General Satisfaction
20. Do you generally feel satisfied with the job overall?	General Satisfaction
1. Are you generally satisfied with your salary for the job you do?	Payment and Incentives
8. Are you happy with the benefits offered by the company?	Payment and Incentives
17. Do you feel adequately compensated/paid for the amount of work you do?	Payment and Incentives
18. Are you generally satisfied with the rewards and recognition for your efforts?	Payment and Incentives
19. How satisfied are you with the incentives provided by the company/institution?	Payment and Incentives
3. Are you happy with the amount of work you must do?	Work Environment
4. Do you feel respected and appreciated by senior management?	Work Environment
9. How would you rate the working environment at the company?	Work Environment
12. How would you rate the communication and feedback provided by senior management?	Work Environment
13. Are you happy with the social interaction you have in the workplace?	Work Environment

14. How do you rate the recognition you receive from your peers?	Work Environment
15. Are you satisfied with the level of autonomy you have on the job?	Work Environment
16. Does the company/institution provide sufficient support for you to succeed?	Work Environment
5. Are you satisfied with the job duties assigned to you?	Workload and Career
7. Are you satisfied with the training and development opportunities offered by the company?	Workload and Career
10. Do you feel as if your skills are being optimally utilized in your current job?	Workload and Career
11. Are you satisfied with the career prospects offered by the company?	Workload and Career

Table 6. Rotated Component Matrix[a]

	Component			
	1. Payment and Incentives	**2. Work Environment**	**3. General Satisfaction**	**4. Workload and Career**
Q18	0,762			
Q8	0,763			
Q19	0,782			
Q17	0,829			
Q1	0,830			
Q3		0,480		
Q15		0,541		
Q13		0,615		
Q12		0,631		
Q16		0,666		
Q4		0,677		
Q14		0,819		
Q9		0,837		
Q2			0,656	
Q6			0,792	
Q20			0,832	
Q11				0,480
Q10				0,491
Q5				0,588
Q7				0,718

Extraction Method: Principal Component Analysis.
Rotation Method: Equamax with Kaiser Normalization.
a. Rotation converged in 15 iterations.

5. FINDINGS

This research investigates the reliability of ChatGPT by correlating it with the Minnesota Job Satisfaction Questionnaire (MSQ). Pearson coefficient correlation method was used to analyse the data. The results of the analysis showed that there was a significant correlation between the two questionnaires at the levels mentioned in Tables 6 and 7.

Table 7. The Correlation of Two Questionnaires

Correlations			
		ChatGPT General	**MSQ General**
ChatGPT General	**Pearson Correlation**	1	,818**
	Sig. (2-tailed)		,000
	N	91	91
MSQ General	**Pearson Correlation**	,818**	1
	Sig. (2-tailed)	,000	
	N	91	91
**. Correlation is significant at the 0.01 level (2-tailed).			

The correlation between the ChatGPT questionnaire and MSQ was P P<0,01, which proves that there is a strong correlation between them. This indicates that the questions in the ChatGPT questionnaire might be highly reliable in measuring job satisfaction.

6. RESULTS AND SUGGESTIONS

The correlation between the ChatGPT questionnaire and MSQ in general is (P<0.01), which proves that there is a strong relationship between them. This also indicates that the questions in this questionnaire might be highly reliable in measuring job satisfaction. The results from each individual dimension and KMO analysis shows that the questions are also valid in measuring job satisfaction.

One of the most difficult problems of conveying research in organizations is to find accurate and flexible questionnaires to assess and gather accurate information about the human resources of the organizations. If used with caution, ChatGPT might offer a practical solution for researchers and HR managers in the field. Moreover, with recent developments in artificial intelligence technology, these applications might be used automatically with little effort by HR managers. This will reduce their costs and save time in terms of research and analysis.

7. LIMITATIONS AND FURTHER RESEARCH SUGGESTIONS

Highlighting the novelty of our research underlines a crucial milestone that has the potential to bring about substantial transformations in the scholarly domain. Artificial intelligence technologies, such as ChatGPT, have the potential to enhance the accessibility and efficiency of scale development procedures, hence facilitating researchers' innovation. This study exemplifies noteworthy advancements in the area, with the potential to catalyse significant changes in academic discourse as this methodology gains broader adoption.

Significantly, based on our current knowledge, there is a lack of prior research in this field that shares similarities with our particular approach and technique. The originality of this work is emphasized by the lack of similar studies, which highlights its innovative character and ability to stimulate new areas of investigation.

Although we had statistically good results from the ChatGPT questionnaire, the findings of this study should be interpreted with caution. The correlation between the ChatGPT questionnaire and the MSQ does not necessarily imply causation. Further research is needed to investigate the causal relationship between the two measures. Furthermore, it is crucial to remember that the study sample size was modest and constrained to a modestly sized universe. Further studies and experiments are required to confirm the results.

REFERENCES

Aamodt, S. P. (2014). *Industrial/organizational psychology: An applied approach* (6th ed.). Wadsworth, Cengage Learning.

Armengol-Estap'e, J., Bonet, O.D., & Melero, M. (2021). On the Multilingual Capabilities of Very Large-Scale English Language Models. *ArXiv, abs/2108.13349*.

Babbie, E. R. (2016). *The basics of social research* (7th ed.). Cengage Learning.

Bakker, A. B., & Demerouti, E. (2017). Job demands-resources theory: Taking stock and looking forward. *Journal of Occupational Health Psychology, 22*(3), 273-285.

ChatGPT (2020). *ChatGPT questionnaire,* https://www.chatgpt.com/questionnaire (Accessed: 12nd February 2023)

Chatman, J. A., & Barsade, S. G. (1995). The influence of organizational culture on employee satisfaction and motivation. *Journal of Applied Psychology, 80*(3), 330-339.

Cheng, K., Li, Z., Li, C., Xie, R., Guo, Q., He, Y., & Wu, H. (2023). The potential of GPT-4 as an AI-powered virtual assistant for surgeons specialized in joint arthroplasty. Annals of Biomedical Engineering, 51, 1366-1370. https://doi.org/10.1007/s10439-023-03207-z

Clark, A. E., & Senik, C. (2016). Who compares to whom? The anatomy of income comparisons in Europe. *Economic Journal, 126*(598), 459-488.

Demerouti, E., Bakker, A. B., & Schaufeli, W. B. (2005). Burnout and work engagement: A two-sample test of the job demands-resources model. *Journal of Applied Psychology, 90*(2), 138-148.

Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). SAGE Publications.

Hackman, J. R., & Oldham, G. R. (1975a). Development of the job diagnostic survey. *Journal of Applied Psychology, 60*(2), 159-170.

Hackman, J. R., & Oldham, G. R. (1976b). Motivation through the design of work: Test of a theory. *Organizational Behavior and Human Performance, 16*(2), 250-279.

Hakanen, J. J., & Schaufeli, W. B. (2012). Do burnout and work engagement predict depressive symptoms and life satisfaction? A three-wave seven-year prospective study. *Journal of Affective Disorders, 141*(2-3), 415-424.

Harter, J. K., Schmidt, F. L., & Hayes, T. L. (2002). Business-unit-level relationship between employee satisfaction, employee engagement, and business outcomes: A meta-analysis. *Journal of Applied Psychology, 87*(2), 268-279. https://doi.org/10.1037/0021-9010.87.2.268

Hernandez, R. (2020). *ChatGPT: A revolutionary technology for automating customer service,* https://www.entrepreneur.com/article/357082 (Accessed: 1st March 2023).

Hilger, N., & Beauducel, A. (2017). Parallel-forms reliability. In V. Zeigler-Hill, & T. Shackelford (Eds.), *Encyclopedia of personality and individual differences* (pp. 1337-1345). Springer.

Ironson, G. H., Smith, P. C., Brannick, M. T., Gibson, W. M., & Paul, K. B. (1989). Construction of a job in general scale: A comparison of global, composite, and specific measures. *Journal of Applied Psychology, 74*(2), 193-200.

Judge, T. A., Bono, J. E., Thoresen, C. J., & Patton, G. K. (2001). The job satisfaction-job performance relationship: A qualitative and quantitative review. *Psychological Bulletin, 127*(3), 376-407.

Kunin, T. (1955). The construction of a new type of attitude measure. *Personnel Psychology, 8*(1), 65-77.

Locke, E. A. (1976). The nature and causes of job satisfaction. In M. D. Dunnette (Ed.), *Handbook of industrial and organizational psychology* (pp. 1297-1349). Chicago, IL: Rand McNally.

Luthans, F., & Youssef-Morgan, C. M. (2017). Psychological capital: An evidence-based positive approach. *Annual Review of Organizational Psychology and Organizational Behavior, 4,* 339-366.

Mayo, E. (1933). *The human problems of an industrial civilization.* Macmillan.

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (1955). A proposal for the Dartmouth summer research project on artificial intelligence. *AI Magazine, 27*(4), 12-14.

Miotto, Marilù & Rossberg, Nicola & Kleinberg, Bennett. (2022). Who is GPT-3? An exploration of personality, values and demographics. *arXiv preprint arXiv:2209.14338.* https://doi.org/10.48550/arXiv.2209.14338

Newell, A., & Simon, H. A. (1961). GPS, A program that simulates human thought. In E. A. Feigenbaum, & J. Feldman (Eds.), *Computers and thought* (pp. 279-293). McGraw.

Nguyen, C. M. A., & Ha, M.-T. (2023). The interplay between internal communication, employee engagement, job satisfaction, and employee loyalty in higher education institutions in Vietnam. *Humanities and Social Sciences Communications, 10*, 329. https://doi.org/10.1057/s41599-023-01806-8

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.

OpenAI. (2023). *GPT-3: The third generation of OpenAI's generative pre-trained transformer,* https://openai.com/blog/gpt-3/ (Accessed: 9th February 2023).

OpenAI. (2023). *OpenAI and Microsoft extend partnership,* https://openai.com/blog/openai-and-microsoft-extend-partnership/ (Accessed: 28th April 2023).

Ostroff, C. (1992). The relationship between satisfaction, attitudes, and performance: An organizational level analysis. *Journal of Applied Psychology, 77*(6), 963-974.

Oswald, A. J., Proto, E., & Sgroi, D. (2015). Happiness and productivity. *Journal of Labor Economics, 33*(4), 789-822.

Pereira, F., & Schabes, Y. (1992). Inside-outside reestimation from partially bracketed corpora. *Proceedings of the 30th annual meeting on Association for Computational Linguistics* (pp. 128-135). Association for Computational Linguistics.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE, 77*(2), 257-286.

Dale, R. (2021). Gpt-3: What's it good for?. *Natural Language Engineering, 27*(1), 113-118.

Rosen, C. C., Levy, P. E., & Hall, R. J. (2016). Placing perceptions of politics in the context of the feedback environment, employee attitudes, and job performance. *Journal of Applied Psychology, 101*(10), 1435-1448.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature, 323*(6088), 533-536.

Saks, A. M. (2019). Antecedents and consequences of employee engagement revisited. *Journal of Organizational Effectiveness: People and Performance, 6*(1), 19-38.

Smith, P. C., Kendall, L. M., & Hulin, C. L. (1969). *The measurement of satisfaction in work and retirement: A strategy for the study of attitudes*. Rand McNally.

Spector, P. E. (1985). Measurement of human service staff satisfaction: Development of the Job Satisfaction Questionnaire. *American Journal of Community Psychology, 13*(5), 693-713.

Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient Alpha and internal consistency. *Journal of Personality Assessment, 80*(1), 99-103.

Sulea, C., Haţegan, I. G., Maricutoiu, L. P., & Schaufeli, W. B. (2015). Work engagement as mediator between job characteristics and positive and negative extra-role behaviors. *Career Development International, 20*(6), 590-609.

Taylor, F. W. (1911). *Principles of scientific management.* Harper & Brothers.

Tims, M., Bakker, A. B., & Derks, D. (2013). The impact of job crafting on job demands, job resources, and well-being. *Journal of Occupational Health Psychology, 18*(2), 230-240.

Weber, M. (1947). *The theory of social and economic organization.* Free Press.

Weiss, D. J., Dawis, R. V., England, G. W., & Lofquist, L. H. (1967). *Manual for the minnesota satisfaction questionnaire*. Minnesota Studies in Vocational Rehabilitation, XXII. University of Minnesota.

Declaration of Contribution Rate: The first author contributed 60%, and the other authors contributed 20% each.
Declaration of Support and Appreciation: The research did not receive any support from any institution or organisation.
Declaration of Conflict: The authors declare that there is no conflict of interest.
In this study, the rules stated in the **"Higher Education Institutions Scientific Research and Publication Ethics Directive"** were followed.
This article has been screened with **similarity** detection software.