

Bias, Halo Effect and Horn Effect: A Systematic Literature Review

Norzalina Noor¹, Sukor Beram², Fanny Khoo Chee Yuet³,
Kumaran Gengatharan⁴, Mohamad Syafiq Mohamad Rasidi⁵

^{1,5}Faculty of Language and Communication, Sultan Idris Education University, Perak, Malaysia, ²Department of Economy, Perak Matriculation College, Malaysia, ³Faculty of Management and Economic, Sultan Idris Education University, Perak, Malaysia, ⁴Institut Pendidikan Guru, Kampus Sultan Abdul Halim, Sg. Petani, Kedah, Malaysia.

Corresponding Author Email: norzalina@fbk.upsi.edu.my

Abstract

A bias is a prejudice that generally unfavorably supports or opposes one thing, person, or organization over another. Individuals, groups, and institutions can all have biases, which can have bad or good outcomes, and most of the focus surrounding the Halo and the Horn effect. The present article set out to analyze the existing literature on the bias, the Halo effect, and the Horn effect. Guided by the PRISMA Statement (Preferred Reporting Items for Systematic reviews and Meta-Analyses) review method, a systematic review of the Scopus databases, ERIC, JSTOR, and Emerald Insight identified 17 linked studies.

Further review of these articles resulted in six main themes: Teachers and teaching, students' tasks, Management, crime, Psychology, and raters. These six themes further produced a total of 30 sub-themes. Several recommendations are highlighted related to conducting more mixed-method studies, having a specific and standard systematic review method for guide Research Synthesis in bias, the Halo and the Horn effect, and practicing complementary searching techniques such as citation tracking, reference searching, snowballing, and contacting experts.

Keyword: Systematic Literature Review, Bias, Halo, Horn and PRISMA.

Introduction

In interpretation or simply giving a perception of a product or service, the individual will not be separated by making an assessment. Therefore, the assessment made cannot escape bias. However, bias towards a product or service can be reduced if the individual clearly understands the meaning of bias and the elements. Many studies have been done on this issue of bias in evaluation. Moreover, we are often with the concept of the halo effect and the Horn effect in bias. This study will focus on these three keywords in exploring and researching in more depth about the research that revealed these three keywords in their study.

Biases

Assimilation-by-association is a term used to describe this bias (Steglich & Knecht, 2014). According to research, teacher-perceived friendships result in an overestimation of similarity among students considered to be friends. Externalising behaviour problems and academic orientation (Gest, 2006), academic expectations and study efforts (Steglich & Knecht, 2014) have all been linked to this bias in teacher judgment (Marucci et al., 2021)

Cognitive biases may be accountable for some misunderstandings. These cognitive biases can have a substantial influence on how individuals decide. Some of the nine cognitive biases or heuristics described in their framework were visible in decision making, whereas others were absent or difficult to recognise (Matlock et al., 2017). Furthermore, cognitive biases can affect judgments based on such impressions. This possible difficulty, namely cognitive biases that influence judgment under predetermined and limited information situations, is one that instructors may encounter (Gweon et al., 2017).

On top of that, ratings also could be biased in favour of or against specific groups (e.g., gender bias). Then there is the possibility of over-generalisation. A general impression of the teacher may overpower performance disparities in other categories (i.e., ratings), which may be highly correlated across categories (Sonnert et al., 2018). Consequently, students may be placed in suboptimal educational tracks based on biased scoring (Jansen et al., 2021). People may be clueless about their unconscious biases, so it is critical to warn them about their proclivity for bias in decision-making, as awareness of bias can help mitigate some of the negative consequences of decision-making (Ehrlinger et al., 2005)

An initiative has been taken to minimise the raters' conflicting attitudes and biases that they might have brought to their assessment by giving them assessment criteria and performance samples and asking them to take part in the training program. A rubric can be used to assess a specific set of assessment criteria. Range restriction, such as the halo effect, is characterised by this overly consistent usage of the rubric (Myford & Wolfe, 2004). Furthermore, the halo effect was as prevalent among raters who read that the performance score on technical competence should not influence the rating of interpersonal abilities as it was among those who did not read the explanation of the bias (Cantarelli et al., 2020).

Halo effect

Thorndike initially noticed the halo effect when he discovered that supervisors could not judge their subordinates separately on various character personalities. The halo effect is a cognitive bias that causes people to see another person's qualities in a way that matches their past impressions of other characteristics. This phenomenon is known as the halo effect, meaning that positive qualities are more likely to be ascribed to individuals who display positive qualities in other domains (Thorndike, 1920).

The term "halo" comes from the Greek word for a ring of light that surrounds the sun or moon. The halo effect allows for distilling perceptions down to a uniform image and avoiding cognitive dissonance. The halo effect follows a tendency to consistent assessment. People attempt to create a consistent and coordinated image of a person or object they perceive (Nufer, 2019). The halo effect works in two directions: On one hand, positive information produces a positive assessment of attributes, while similarly, negative information tends to result in a rather negative judgment of attributes (Graf & Unkelbach, 2016). On the other hand, the halo effects occur when the characteristics to be assessed are

ambiguous or complicated to ascertain but may nevertheless be distinctive enough to impact how discernible characteristics are rated (Landy & Sigall, 1974).

When interpreting behaviours observed in others, people frequently underestimate the role of context and overestimate personality-based influences, known as the fundamental attribution fallacy (Ross, 1977). The halo effect's documentation in scenarios when an educator evaluates learners has yet to be investigated (Gweon et al., 2017). However, Jansen et al (2021) stated that spelling problems caused halo effects on analytic rating scales. These halo effects could be reduced by using a prompt. The quantity of spelling errors, for instance, has been shown in empirical investigations to influence teacher perceptions of other independent features, a phenomenon known as the halo effect. The term halo effect refers to the impact of one characteristic on the evaluation of another. Halo effects can have severe consequences for students as they can lead to errors in other aspects of writing, such as content, organisation, argumentation, or essay structure, and thus potentially trigger students to lose motivation and confidence.

The halo effect was helpful to demonstrate to students how easily their standard information processing could be biased. If a person's Halo (a person's most important quality) is reflected in another trait of the same individual, the effect occurs. When both attributes are virtually independent of each other, this could still happen. As a result, a student who is thought attractive by a teacher may also be deemed intelligent, even if he or she is not. The halo effect has been found to significantly affect teachers' evaluations of student achievement (Dompnier et al., 2006).

The halo effect is the perception of a company's success or failure based on its reputation. Companies can be the object of a 'halo effect' when their reputation is high. A brand's excellent reputation has a significant impact on the overall perception of product quality. The halo effect affects how people perceive the quality of products and corporate culture. The product may cause them to associate the image of the country of origin with unfamiliar product features. In marketing communications, celebrities are frequently used as endorsers to enable a product to gain an advantage from a halo effect. Especially when launching new products, a company stands to benefit from the halo effect. The image of a well-established brand is transferred to a newly introduced product bearing the same brand name, thereby motivating consumers to purchase (Nufer, 2019).

Halo effects were evident in health as well. The participants appeared to impart a halo effect in two places which are appeared in the vicinity of the clinician and the technology itself (Matlock et al., 2017). The halo effect refers to positive behaviours such as seeing attractive individuals as more successful and popular. The reverse Halo effect occurs when perception has a negative meaning, such as considering attractive individuals to have undesirable attributes (or Horn effect) (Marucci et al., 2021). For example, attractiveness can generate a "halo" effect, in which others believe that more attractive individuals are more interpersonally and professionally proficient. This halo effect is most likely related to the projection of interpersonal desires: people want to form ties with attractive individuals (e.g., as romantic partners or friends). Positive interpersonal skill evaluations of attractive individuals reflect their motivations for engaging in such relationships. Non-attractiveness, on the other hand, causes a "Horn" effect, suggesting that encounters with non-attractive individuals may make negative information more salient than positive information (Xu et al., 2020). This explanation is in line with the halo effect (Nisbett & Wilson, 1977), which postulates that global, positive, person-centred evaluations of an individual (in their case,

someone's level of adjustment) render positive evaluations of that person's other attributes (e.g., creativity) (Sutu et al., 2019).

According to existing research, the efforts to maintain cognitive consistency in ratings of all measures aimed by effect felt toward the target (the so-called halo effect) are far more significant in multiple-rater assessments. (Kahneman, 2011; Thorndike, 1920). This unintentional but powerful cognitive bias leads to high interdependencies among all rated qualities, even conceptually unrelated ones (Brown et al., 2017). The halo effect refers to "how judgments about some aspects of an object may influence how other aspects of an object are judged" (Chartered Institute of Personnel and Development (CIPD), 2015, p. 24). Halo effects were highlighted as a specific issue that would require thorough attention before ratings of this type could function properly (Sonnert et al., 2018)

Horn Effect

A 'reverse halo effect' is accountable for these stigmatising assessments. Halo effects are a cognitive bias that affects how people create impressions and decide (Thorndike, 1920). A positive halo effect happens when an object is assessed to have positive attributes in other domains (e.g., attractiveness) due to having a favourable attribute in one area (e.g., intelligence). As a result, a negative or reverse halo effect (also known as the "devil effect" or "Horns effect") might occur. Forgas and Laham (2017) assume someone has negative attributes determined by another, unrelated undesirable or stigmatised characteristic. This characteristic can include a learning disability (Shifrer, 2013), unattractiveness (Gibson & Gore, 2015), and negative behaviours, such as Children's oppositionist (Jackson & King, 2004) and steroid use in athletes (Chantal et al., 2013).

Methodology

In this section, the method used to retrieve articles related to the Halo effect and Horn Effect as an integral of the biases effect. The reviewers used the method, namely PRISMA, which includes resources used to run the systematic review process, data abstraction, and analysis.

PRISMA

In line with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA; see Figure 1), the review methodology involved four steps - identification, screening, eligibility and inclusion. As part of the identification step, many electronic databases were used to conduct Bias, Halo, and Horn effect keywords (Figure 1). PRISMA statement 2020 provides updated reporting guidance for systematic reviews that reflect advances in identifying, selecting, appraising, and synthesizing studies (Page et al., 2021). It offers three unique advantages, which are, 1) defining straightforward research questions that permit systematic research, (2) identify inclusion and exclusion criteria (Table 1 and Table 3) it attempts to examine an extensive database of scientific literature in a defined time (Figure 1). In addition, the PRISMA allows for a rigorous search of terms related to bias, the Halo effect and the Horn effect.

Resources

The review drew on four of the most widely used journal databases, Scopus, Emerald Insight, Eric, and JSTOR. Scopus is a robust database used in the review as one of the largest abstract and citation databases for peer-reviewed literature. With over 22,800 journals from 5000

publishers worldwide, Scopus covers the broadest range of abstract areas, including environmental sciences, social sciences, and biological studies. ERIC, the Education Resource Information Center, provides access to scholarly and research-based materials on education. Eric indexes a broad range of journal sources, including journal articles, research reports, curriculum and instruction guides, conference papers, dissertations and theses, and books dating back to 1966. Discover journals, books, and case studies with Emerald Insight. Browse over 2,500 eBook titles and a portfolio of over 300 journals written by renowned academics and industry leaders from over 150 countries. The databases provide access to titles from various subject areas, ranging from management and finance to criminology and social media. JSTOR is a not-for-profit organisation dedicated to providing an intellectually curious community with access to a digital library. JSTOR contains over 12 million academic journal articles, books, and primary sources spanning 75 disciplines. JSTOR is one of the most trusted sources of academic content globally, with millions of users accessing it for research, teaching, and learning.

Eligibility and Exclusion Criteria

Numerous criteria for eligibility and exclusion are identified. First, only articles in peer-reviewed journals containing empirical data are considered concerning literature type, which means that review articles, book series, books, chapters in books, and conference proceedings are all excluded. Second, to avoid misunderstandings and complexity in translation, the search efforts excluded non-English publications and concentrated exclusively on English-language articles. Thirdly, ten years (between 2017 and 2021) are selected to observe the evolution of research and related publications. Fourth, due to the review process's emphasis on the bias, the Halo effect, and the Horn effect, finally, in keeping with the journal's objective of assessing bias in assessment, all journals containing this keyword in one's measurement statistic are excluded.

Systematic Review Process

Four stages were involved in the systematic review process that was performed in July 2021. The first phase identified keywords used for the search process. Next, the terms similar and related bias were used (Table 2). At this stage, two duplicated articles were removed following a thorough screening. The screening was the second stage. At this point, 621 of the 653 articles eligible for review have been removed. The third stage is eligibility, during which all articles were accessed in their entirety. After careful examination, a total of five articles were excluded because they focused exclusively on systematic reviews (2), review articles (1), articles based on guideline development with no mention of the method used (1), articles on statistical issues (1), and articles that did not measure bias in evaluation (1). The final stage of review yielded 17 articles for qualitative analysis (see Fig. 1).

Data Abstraction and Analysis

The rest of the articles were reviewed and analysed. Efforts were focused on studies that addressed the planned questions. The data was extracted by first reading the abstracts and then the full articles (in-depth) to identify relevant themes and sub-themes. A qualitative analysis was conducted using content analysis to identify themes relating to bias, the Halo effect, and the Horn effect concerning behaviour. The authors then organised sub-themes

around the typological themes. A flow chart illustrates how PRISMA was used to conduct a qualitative synthesis of published bias, Halo, and Horn effect studies between 2017 and 2021.

Table 1

The inclusion and exclusion criteria

Criterion	Eligibility	Exclusion
Literature type	Journal (research articles)	Journals (systematic review), book, book series, chapter in a book, conference proceeding, research report, document, Early cite article, case study, expert briefing, executive summary, conference review, letter, erratum
Language	English	Non-English
Timeline	Between 2017 and 2021	<2017

Table 2

The Search string used for the systematic review process

Databases	Keywords used
Scopus Screening	TITLE-ABS-KEY ("Halo effect" OR "Horn effect" AND bias) AND (LIMIT-TO (OA, "all")) AND (LIMIT-TO (PUBYEAR , 2021) OR LIMIT-TO (PUBYEAR , 2020) OR LIMIT-TO (PUBYEAR , 2019) OR LIMIT-TO (PUBYEAR , 2018) OR LIMIT-TO (PUBYEAR , 2017)) AND (LIMIT-TO (DOCTYPE , "ar")) AND (LIMIT-TO (LANGUAGE , "English")) AND (LIMIT-TO (SRCTYPE , "j"))
ERIC Screening	TITLE-ABS-KEY ("Halo" OR "Horn" AND "bias") Limiter: Full text Expander: any related word Search Modes: Find all my search terms.
JSTOR Screening	All (Halo effect) OR (horn effect) AND bias Academic Content: Journals Date: 2017-2021
Emerald Insight Screening	TITLE-KEYWORD ("Halo effect" or "horn effect" or bias) Content-Type: Article Year: 2017-2021

Result

The review resulted in six main themes and 30 sub-themes related to bias, Halo, and Horn effect studies. The six main themes are teachers (six subthemes), students' task (six subthemes), management (eight subthemes), crime (two subthemes), psychology (three subthemes), and raters (five subthemes) (Figure 2). The results provided a comprehensive analysis of the current studies about bias, effect, and horn effect influencing an individual's judgment and decision-making.

A total of four studies out of five studies focused on Bias and Halo effect (Jones & Bergin, 2019; Matlock et al., 2017; Sonnert et al., 2018; Sutu et al., 2019), and one study focused on the Halo effect only (Sheppard, 2019) in United States of Amerika (USA). Two studies out of three focused on Bias, Halo effect and Horn effect (Marono & Bartels, 2020; Xu et al., 2020), and only one study focus on bias and Halo effects (Brown et al., 2017) in the United Kingdom (UK), three studies focused on bias and Halo effect (Behrmann, 2019; Jansen et al., 2021; Nufer, 2019) in German, five countries focused on the exact keywords which are Bias and Halo effect in Brazil (Rego et al., 2018), Turkey (Yilmaz, 2017), Italy (Cantarelli et al., 2020), Iran (Bijani, 2018) and Korea (Gweon et al., 2017). Finally, one study concentrated on the Dutch's bias, Halo effect and Horn effect (Marucci et al., 2021) (Table 3).

Furthermore, four studies used a qualitative approach (2 studies used interviews and one used observation). Three studies employed mixed methods (quantitative & qualitative) design, one of which is empirical research. Both studies involving an experimental approached. The remaining studies (10) used a quantitative approach. Regarding years published, two articles were published in 2021, three were published in 2020, five were published in 2019, three were published in 2018, and four were published in 2017.

Bias, Halo effect and Horn effect studies

This section focused on the main Bias, Halo effect and Horn effect applied by the researcher in their studies in six main themes such as teachers and teaching (six subthemes), student s' task (six subthemes), management (8 subthemes), crime (2 subthemes), psychology (three subthemes), and raters (five subthemes).

Teachers and Teaching

A total of six out of 17 studies concentrates on teachers and teaching. The most common bias, Halo and Horn effect studies, are teachers' assessment (5 studies), while two studies found that teachers' attitude influenced bias. There are four studies focused on only one subtheme: teacher attunement, teacher education, and teaching practice towards bias. Teachers' assessment and judgment are some of the practices that need teachers to measure students' tasks. The problem with this type of bias is that it often occurs outside of teachers' conscious grading process.

Four studies out of five measure student tasks were biased. For example, Jansen et al. (2021) used 51 pre-service teachers to assess four argumentative essays on a holistic scale before they assessed seven individual characteristics on analytic scales. Meanwhile, Sheppard (2019) used teachers to evaluate 122 writing assignments in five rubric dimensions. Whereas Yilmaz (2017) used 24 teachers to rate the diagnostic tree prepared by the teachers' candidates and Gweon et al (2017) used teacher to evaluate 22 students worked on one collaborative project in three-phase. Instead of One study not measure students' tasks but concentrated on 56 teachers' judgment on their ability to recognise bullies, victims and

prosocial students, and at the same time, this study also measures teacher attunement (Marucci et al., 2021).

This study also found a second subtheme that is the teachers' attitude. This subtheme involves two studies: teachers' attitudes towards the bias, as which teachers' awareness is essential for stimulated to minimise their effect (Marucci et al., 2021). Teachers' attitude change depended on both cognitive and affective reactions. For the teacher education subtheme, the study also showed that the model of teacher change induced a considerable number of halo effects on teacher candidates (Behrmann, 2019). Vice versa of that practice, only one study used 85 undergraduate students to evaluate pre-college mathematics teachers through video (Sonnert et al., 2018). Furthermore, one study measured biases in eight teaching practices among 421 principals in an authentic teacher evaluation system used across schools in over 10% of observations.

Students' Task.

A teacher might get to know their students better, but this process is still very complex; all kinds of biases such as those mentioned above could get in the way. Under students' tasks sub-theme, a total of four studies out of 17 to test a Bias, Halo effect and Horn effect used rubrics they developed Sheppard (2019) used five dimensions of a rubric to measure the degree to which scores correlated with scores for each dimension. Yilmaz (2017) rate the diagnostic tree prepared by the teacher candidates that the criteria were gathered opinions from both experts and teacher candidates to make sure the criteria were intelligible. Jones & Bergin (2019) make multiple observations using a rubric based on INTASC standards by trained evaluators. Instead of other studies, Gweon et al (2017) not using any rubric in their study, but they suggest using provided guidelines that would enable instructors to be less prone to psychological biases such as the Halo effect.

A total of two out of 17 studies measure writing tasks. Writing essays asserts the author's opinion on a topic, whether academic, literary, or humorous. There are various approaches to essay writing and millions of different topics to choose from as themes. However, good essay writing tends to have a very similar framework. For example, Jansen et al (2021) measure argumentative essay writing that evaluates the text from the aspect of spelling errors as an integral of the criteria in the rubric used, while Sheppard (2019) uses five dimensions in measuring writing assignments to see the relationship of each of these dimensions to the existence of the Halo effect.

Furthermore, from the group work sub-theme, there are two studies involved. Group work refers to individuals working together and carrying out tasks shoulder to shoulder to achieved team goals. A team consists of a combination of several individuals who have interdependent competencies in terms of ability, expertise, skills, knowledge, accountability and commitment to team performance and are willing to perform tasks together. Yilmaz (2017) investigated the effect of rater participating in an evaluation that spanned two categories, namely group work and individual work, and found no Halo effect to group compared to an individual. On the other hand, Gweon et al (2017) stated that rater would be vulnerable to human error, Halo effect and fundamental attribution error because they have limited views on group work due to task distribution and may occur isolated group members.

The last subtheme involving one study for each subtheme was spelling, oral and video. As students, we cannot run away from committing grammatical errors, especially errors in spelling. Spelling errors are often seen as an essential component in writing scoring. However,

Jansen et al (2021) argues that the texts with spelling errors, whether the text is low or high quality in writing, should not be influenced by such spelling, which that spellings are not the cause of the Halo effect evaluation.

Besides, there was only one subtheme that conducted the study on oral assignments. This sub-theme is a fundamental skill that needs to be mastered by every student. Students need to master oral skills first before mastering other language skills such as reading and writing skills. Students who can master oral skills are usually able to master other skills well too. Bijani (2018) states that the researcher observed biased interactions between assessors and test takers to test high language performance ability on an ongoing basis.

Finally, there is one study that looked at video presentation measurement. The video presentation allows students to train themselves to appear bold and not clumsy to speak in public. If in the classroom, this activity is a learning process that we should all go through, and with the training, we can adapt all our knowledge and experience when working later. It can be used as an added value for oneself when entering the world of work. Nevertheless, the study of biases in this video task did not measure the videos produced by the students. Instead, the undergraduate students measured the videos of the pre-college mathematics teachers while teaching. The purpose of this study is to determine whether there is a bias impact in the assessment, as an alternative to engaging experts, which is time consuming and costly.

Management

Management in this study involves the overall administrative process of the organization, which includes the problems of human resource management, information systems management, administration, and development. Management also experiences bias, the Halo effect and the Horn effect, consciously or unconsciously. Eight subthemes in management will be peeled one by one in this study. Under this theme, a total of five subthemes emerged on bias in decision-making. Five studies focused on cognitive bias. Two studies focused on gender biases, and the remaining seven studies representing one subtheme each, namely ideal employee, sport, manager attitudes, aged biases, appearance.

The majorities focus in this study is the decision-making and cognitive bias subtheme. A total of five studies out of 17 were studies bias in decision-making. Rego's (2018) study found the effect of bias on Brazilian managers' attitudes toward older employees and how those attitudes are influenced by their age to explain Human Resources Management decision-making in behavioral intentions. Along with that as well, Matlock et al (2017) found evidence of cognitive bias in decision-making for implantable cardioverter-defibrillators (ICD) implantation. A large number of these biases appear to drive ICD treatment. Due to bias problems in their studies and to produce suitable constructs with high convergent validity to avoid bias, Brown et al (2017) have used a forced-choice stage modelled with Thurstonian item response theory (IRT). The results of this study found that raters' agreement was slightly better in decision-making. In addition, Cantarelli et al (2020) states that debiasing technique interventions can also eliminate the effects of merging and framing and improve decision-making. On the other hand, Bijanis' (2018) study found that decision-makers should not care about the level of expertise of the rater. However, they should set up rater training programs to improve rater consistency and reduce their biases in measurement.

Further, a total of five studies were cognitive biases subtheme. Humans are thinking beings. With these thoughts, man makes his daily decisions. However, without realizing it,

rationally, it turns out that decisions and everything we believe in is influenced by a single psychic factor or known as cognitive biases. Human attention is limited, and we tend to think heuristically or take “shortcuts” from general logic to make decisions. Unbeknownst to us, the “shortcut” leads to cognitive biases. Nufer (2019) states that cognitive bias leads to skewed assessments. Based on the halo effect in Marucci et al (2021) study, cognitive bias will affect teachers’ ability to recognize bullies, victims, and prosocial students. Cantarelli et al (2020) study investigating the effects of various cognitive biases and interventions showed that decision-making relied heavily on patterns of systematic deviations from rationality and debiasing interventions were only occasionally successful in eliminating such deviations.

Furthermore, Behrmann (2019) states that teachers’ daily use of cognitive heuristics can lead to bias in information processing and, thus, unfair student assessment. The application of empirical research principles periodically and instruments retrospectively can help in reducing cognitive bias. However, their findings contradict the assumption that was addressing cognitive engagement and emotional effects influence attitudes. Moreover, Matlocks’ et al (2017) study explains many cognitive biases, and there may be biases that may not be identifiable in their study. Cognitive bias can influence decisions in non-normative ways. Of the nine cognitive biases, they found the effects of framing that most apparent effect, followed by default and Halo effect; some evidence of optimism bias, affect heuristics, state dependence, anchoring and availability bias; and little or no evidence of effective prognosis.

Although some individuals deny gender bias exists in every aspect of society, from the workplace to the political arena. However, that bias is existing. The gender gap may affect our children’s education, the size of the salary we take home, and why women still lag behind men who have specific careers. These statements are evidenced by three studies conducted by the following scholars. Xus’ et al (2020) results suggest that the gender of the provider can influence the response to service providers. However, gender does not play a moderate role in the relationship between attractiveness and interpersonal skills, and customers rate their satisfaction based on employee interpersonal skills, regardless of gender. However, the opposite result occurred in Rego’s (2018) study when it found that female managers preferred older female employees. However, their findings do not support a “double hazard effect” on older female workers. Gender equality tendencies emerged as female managers preferred older female employees compared to younger male employees. Such bias does not appear when the older employee is male. The lack of this study, when this study does not look at scenarios that show older employees as men and younger ones as women, so it is not possible to know whether gender equality bias also affects male managers. In other words, Sonnert et al (2018) examined gender biases in education and found that male students handed out lower evaluations than female students in the rating situation.

In contrast, teacher gender did not significantly influence the ratings. However, this resulted in was the primary influence of baseline assessment as the only source of gender bias. It is unclear how much of their findings are linked to the influence of discipline. Students rater have never been in these teacher’s classes or other rating experimental aspects or parameters. The lack of gender interaction effects between teachers and students is a positive finding. The main effects of gender detected for raters may be relatively easy to control and manage, especially if the raters have a stable gender composition.

The remaining studies represented one subtheme, namely ideal employee, sport, manager attitudes, aged biases, attractiveness, interpersonal skill and self-esteem. Each

employer will hire employees who will fit the company culture and fit the company's mission. In most cases, the quality of employees must match their personalities to be ideal employees. However, the effect of bias can cause employers to choose the employees they want mistakenly. Brown et al (2017) suggest that the mechanism for this improvement is a better differentiation between behaviors in comparative assessment and applied IRT to Multisource feedback (MSF) collected using multidimensional forced-choice response format operations as an effective method of bias prevention. Multisource feedback (MSF) is increasingly used to evaluate trainee performance, with different evaluators fulfilling a vital role in the use of evaluation data (Tariq et al., 2020).

In sport, the Halo effect exists if the effect created by the dominant attribute affects how the attributes of another object or subject are evaluated. However, the Halo effect has hardly been studied in sports-related contexts, although it can substantially understand how sports fans think and behave. The results of the analysis prove that the deviation of fans' in sports perceptions with regards to the very diverse aspects triggered by the success or the failure of their favorite club (Nufer, 2019).

Overall, managers' attitudes and aged biased are interrelated. Rego et al (2018) concluded that managers with more positive attitudes are older and tend to choose older workers than younger workers due to bias. The manager's attitude toward older employees explains the manager's decision, but this relationship is diverse. To better understand relationships, it is necessary to identify attitude profiles, namely the "Quite incompetent but warm" and the "Competent but cool" profiles. As well as a complex interaction between attitudes should be considered. They also included the attitude of competence as the independent variable and warmth as the moderator.

Lastly, Xu et al (2020) explained three appearance subthemes in their study: attractiveness, interpersonal, and self-esteem. The results showed that the effect of the three-way interaction between attractiveness and self-esteem was mediated by the interpersonal skills perceived by the participants. Furthermore, they reaffirmed these findings and showed that attraction positively affected participants' perceptions of interpersonal skills. Furthermore, the relationship between attractiveness and interpersonal skills was moderated by self-esteem, so that the effect was more substantial in response to participants with relatively low self-esteem.

Crime

Judgment bias is a reaction relative to an obscure stimulus, stating the "interpretation" of this stimulus and the "expectation" resulting from the reaction (Roelofs et al., 2016). The knowledge of prior crimes can direct participants' attention to the suspect's crime, artificially enhancing the effect of bias (Maroño & Bartels, 2020). There are studies of crimes theme in two subthemes: Reverse Halo Effect/ debiasing and stigma/ sexual orientation. Two studies out of 17 examined reverse Halo and debiasing. (Maroño & Bartels, 2020) stated that to help isolate the reverse Halo effect, they suggest that people should evaluate individuals in terms of their characteristics (e.g., attractiveness, personality, intelligence) in general and not in an offensive context. If there is a reverse Halo effect, these features will be evaluated negatively. This negative evaluation is because the reverse Halo effect depends on people's belief that all actions committed by criminals are supported by the intent to commit the crime. On the contrary, Cantarelli et al (2020) stated that the debiasing intervention was indeed effective, as the bias effect disappeared when participants were told that people, in decision-making,

might succumb to isomorphic (a one-to-one correspondence (mapping) between two sets that preserves binary relationships between elements of the sets) stress. Additional efforts to design successful debiasing techniques are required, such as implementing motivational strategies with substantial incentives, cognitive strategies with training and technology strategies through external tools such as decision models, decision-making software or group decision making.

In addition, there is only one study that examined stigma and sexual orientation. The stigma associated with sexual crimes can provide negatively biased judgments related to non-sexual offences. Their findings found violation-related actions on which participants should base their judgment were not sexual. Whether this effect is driven by a reverse Halo effect, a schema-based interpretation, or a combination of both, this demonstrates just how damaging the stigma associated with sexual crimes can be.

Psychology

In the theme of psychology, three subthemes are underlying it, namely behaviour (6 studies), personality (2 studies), social psychology (3 studies). Psychological bias is the tendency to make unintentionally erroneous decisions or acts. Psychological bias is the opposite of common sense in a clear and measurable assessment. It can result in missed chances and poor performance. To overcome it, individuals must find strategies to incorporate objectivity into decision-making and allow it more time. Use tools to help evaluate background material systematically, surround people who will challenge our beliefs, and listen carefully and empathetically to their perspectives.

Marucci et al (2021) stated that teachers' evaluation of behavioural qualities differs from how peers are regarded. Students who were perceived as prosocial peers, for example, were regarded as being less aggressive and more popular, liked, and linked. Teachers seem to know that aggression and antisocial behaviour are more valued in popularity by peers, which is contrary to researchers' initial expectations. Notwithstanding, Yilmaz (2017) stated that teacher candidates might rate lower than expected on specific criteria, indicating bias in comprehension and behaviour. In this scenario, the rater's bias is derived from the teacher candidates' professional expertise and experience in terms of evaluation rather than criteria. Apart from this, Many-Facet Rasch provides a tidy method to offence and can provide information on individual behaviour. This tool could help determine the top raters' effect in the Observation of Teaching Practice (Jones & Bergin, 2019)

In contrast, behavioural science shows that assessment is systematically biased in certain situations. From a theoretical point of view, their study highlights behaviour as a secondary distance theory in the discipline of Human Resources (Cantarelli et al., 2020). In addition, Rego et al (2018) argues that age affects behaviour intention, leading to discrimination. Discrimination is defined as actions taken as a result of stereotypes and prejudices. These findings suggest that behaviours, rather than attitudes, are more resistant to bias toward older employees and that this resistance manifests itself even (or mostly) in a downsizing context. Moreover, Xu et al (2020) advise that strategies to assess interpersonal skills and train employees to develop interpersonal skills should be considered improving customer satisfaction with service and behavioural intents. Customers can be influenced by qualities unrelated to the service provider's performance (such as gender) and self-esteem rather than actual employees' behaviour, so using common customer feedback as part of employees' performance should be done with caution.

Along with those opinions, Bijani (2018) found that those raters who had no improvement in their rating behaviour following the training program were among those who had a negative view of the training program and its usefulness in decreasing bias and boosting consistency. Nonetheless, the feedback was found to be quite valuable by the majority of raters, who believed it had a positive impact on their scoring behaviour. As a result, it is essential to note that raters' responses to the feedback question were linked to their receptivity to the feedback and, consequently, the frequency of change in their rating. There are three studies indicated about social psychology. In sport, as defined by social psychology, the Halo effect is a cognitive distortion that involves using well-known attributes of a person to draw judgments about unknown traits or characteristics of that person. The existence of the Halo effect indicates that football fans want the sporting performance of their favourite club to be successful because it enhances their personal life and sense of satisfaction (Nufer, 2019). Whereas in education, when teachers are aware of which students are behaving prosocially, the students can act as a resource for protecting and supporting the victim. Bullying will go unnoticed when it is perpetrated by students whom the teacher considered to have favourable characteristics.

Furthermore, when teachers are considered socially adept, they may overlook victims. Finally, teachers may underestimate students' prosocial behaviour when they are engaging in negative behaviour (Marucci et al., 2021). In addition, Xu et al.'s (2020) study add to the theory of social comparison. They argue that individuals with high self-esteem are more self-confident than those with low self-esteem, and high self-esteem can protect themselves from increased societal comparisons. For example, individuals with low self-esteem may be driven or inspired by attractive people, whereas those with high self-esteem may be less concerned about others' looks.

Two studies examined personality. Only a high association between personality profile normality and perceived creativity (self-rated and informant-rated) was discovered, but not between personality profile normativity and performance-based measures (i.e., behavioural creativity measured in the laboratory, creative achievement, or job creativity). Rater may have implicit theories about creativity, which could have influenced their self-perceptions or informant perceptions on personality and creativity measures (Sutu et al., 2019). On the other hand, the result by Brown et al (2017) study showed a better convergence correlation with an external personality measure related to different personality traits and cultures. As expected, competence self-assessment significantly correlated with self-reported personality, but assessments by external observers correlated weaker with self-reported personalities. Nevertheless, the advantages of either modelling bias or preventing it are more evident than self-assessment, increasing the average validity.

Rater

In the theme of rater, six subthemes are underlying it, namely rater biases (4 studies), rater perception (3 studies), and two studies each for rater training, leniency/ severity, and attribution error. In rater biases, Sonnert et al (2018) suggests that to overcome the rater biases in education, when teachers have little opportunity to receive feedback on their work due to a lack of resources, teachers can use an alternative to direct classroom observation by experts, which opens up assessment systems to give unbiased evaluations of a number of pedagogical factors vital to student learning, as well as comparisons with where other teachers may fall on a scale of the same measures. Teachers can also send digital videos of

their lessons electronically for remote methods. Similarly, Bijani (2018) stated that raters in a face-to-face training session lowered biases more than those in an online training program because they were given feedback after rating. They also believe that a rater who has a positive attitude towards rating feedback is more able to incorporate it into their rating, resulting in more consistency and fewer minor biases in their subsequent ratings.

Conversely, experienced raters, because of their arrogance or overconfidence, did not increase the quality of their judgments as much as new raters by considering aspects of harshness, bias, and consistency. Above all, studies on assessment similarities across all competencies-substantive overlap (actual competencies in the same individual or true Halo) and rater biases (accumulation of competencies in the same individual or illusion circle) Both were at play, with the rater bias having a more significant influence, according to statistical modelling analysis of general method components (Brown et al., 2017). Jones & Bergin (2019) claims that raters' assessments are reliable, consistent with scoring protocols, and free of biases and that raters can assess teacher performance appropriately. Both teachers and their students deserve accurate and fair evaluations. Even with repeated and focused calibration training, these findings occurred when using their rubric. Furthermore, these findings show that if cut scores are employed, observed scores may be biased, resulting in some teachers' erroneous competency.

In rater perception, Xu et al (2020) argue that the underlying mechanisms that explain why more attractive employees tend to receive favorable customer perceptions have not been elucidated by the existing literature. However, their findings suggest that interactions with attractive ones may lead to more positive perceptions of identical behaviors, which elicit more satisfaction and positive behavioral intentions. On top of that, Sutu et al (2019) findings also imply that people may have implicit concepts about creativity that influence their self-perceptions and informant perceptions of personality and creativity measures. Self-perception and other-perception are significantly influenced by personality profile normativeness (i.e., the more well-adjusted someone appears, the more people think they are creative). However, the Halo effect was reduced, and the inconsistency in rating test takers' performance abilities was dramatically decreased by both the raters' pleasant perceptions and attitudes toward training programs (Bijani, 2018).

In severity/leniency, Bijani (2018) stated that upon receiving feedback, the majority of raters changed their evaluation behavior in terms of severity/leniency and bias, and increased consistency, resulting in a genuinely consistent analysis of qualitative and quantitative data. However, some raters are extremely biased. These were the raters who had previously indicated dissatisfaction with the training. It is worth noting that even individuals who were pessimistic about the program or purposefully ignored it could improve their consistency. In one case, a rater was able to reduce his severity and bias to within acceptable limits. It is noteworthy that no raters who had a positive attitude toward feedback obtained a negative outcome.

Furthermore, Gweon et al (2017) claim that the variation of the individual's scores is another result that could indicate a fundamental attribution error during evaluation. For example, suppose the rater assigns a large portion of an individual's success to group performance. In that case, the scores of those in the same group will be similar, resulting in minimal variation. Alternatively, to minimize the attribute errors, Jansen et al (2021) argued that the Halo effect could be decreased if the rater focused more on the assessment's

targeted attributes. A prompt to evaluate each analytic scale as a separate, different attribute and independent category could encourage this.

It is necessary to enhance raters to address the biases that affect training. Marucci et al (2021) suggested that teacher training can be improved by concentrating not just on teaching and students' cognitive development, but also on their socio-emotional development. In addition, training programs effectively change people's attitudes, perceptions, and evaluations of them. Therefore, these training programs almost certainly reduce their intensity and biases and increase their consistency. Furthermore, educating raters about performance evaluation objectives in training programs will reduce the Halo effect (Bijani, 2018).

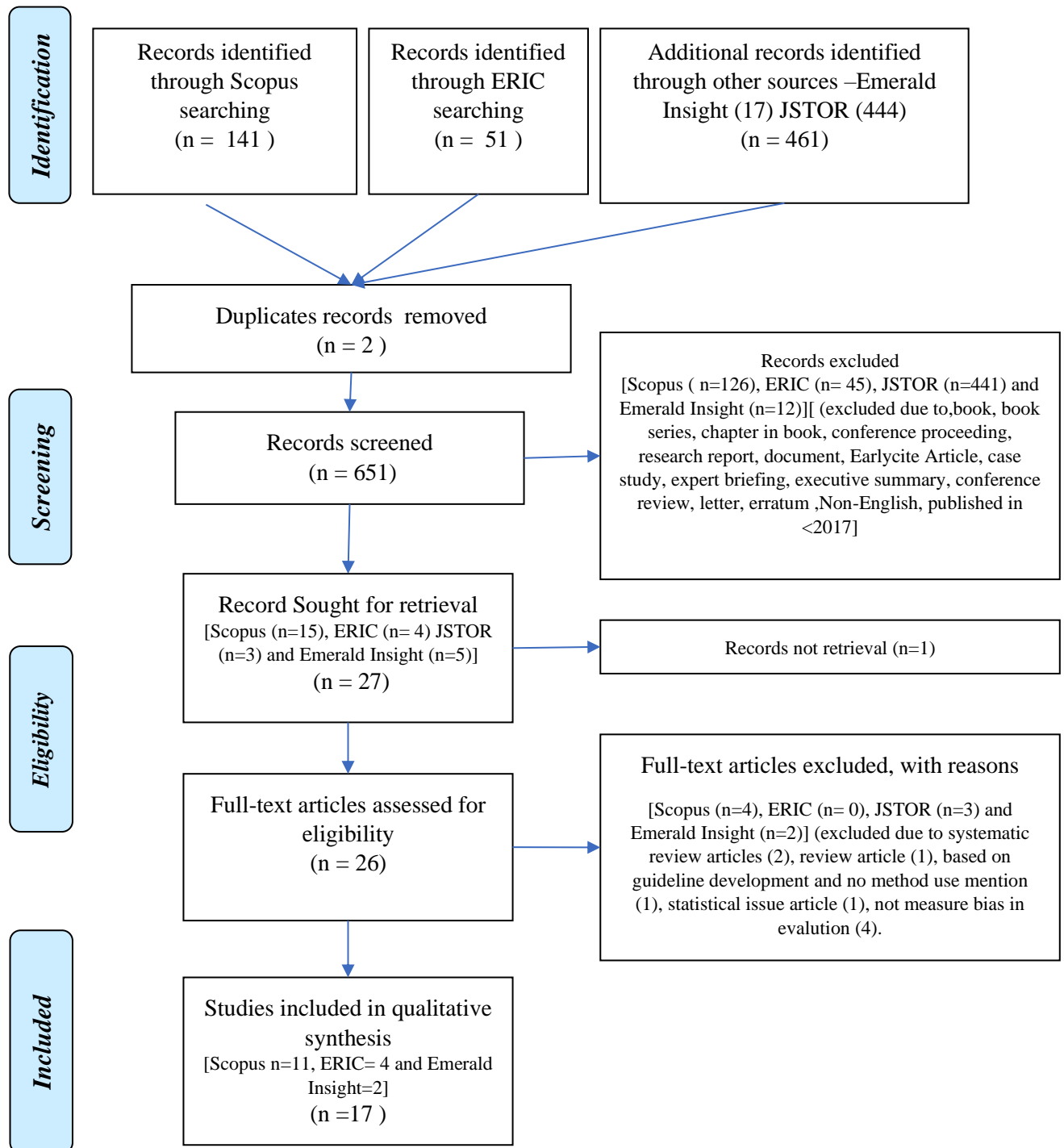


Figure 1: The flow diagram of the study (Identification of new studies via databases and registers)

(Adapted from Page et al (2021).

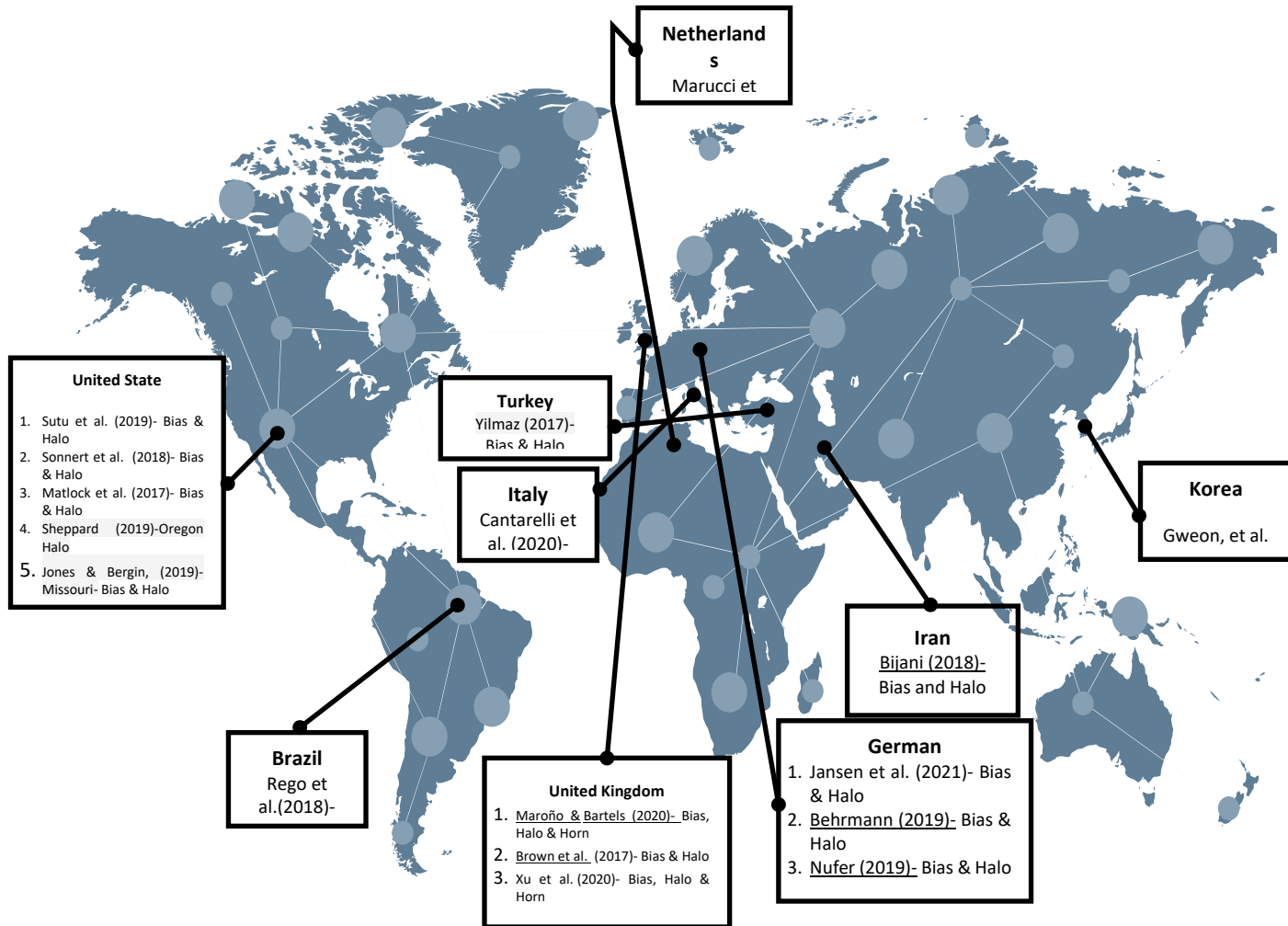


Figure 2: Bias, Halo effect and Horn effect research area of selected studies

Table 3

The findings

Author/Countries	Main study design	Teaching						Students' Task						Management						Crimine		Psychology			Rater								
		AT	TT	TE	TJ	TS	TP	RM	W	S	O	V	PR	GW	C	DM	ID	S	GB	MA	AB	A	RH	SS	P	SP	B	RT	RP	RB	LR	RE	
(Marucci et al., 2021)- Netherlands	Quan	☑	☑			☑									☑										☑	☑	☑						
(Jansen et al., 2021)- German	Quan					☑		☑	☑																								☑
(Maroño & Bartels, 2020)- LondonUK	Quan																						☑	☑									
(Cantarelli et al., 2020)- Italy	Quan														☑	☑							☑			☑							
(Sutu et al., 2019)- USA	Quan																							☑				☑					

Discussion

The purpose of this study was to comprehensively review the existing literature on the bias, halo effect, and Horn effect in evaluation. Evaluation is judging or calculating the quality, significance, quantity, or value of something. Several strategies must be adopted before implementing measures to mitigate the effects of bias, Halo, and Horn, and capitalise on potential opportunities to avoid them. A thorough review of four databases yielded 17 articles on rater bias, the Halo effect, and the Horn effect. The findings indicated that bias, the Halo effect, and the Horn effect had been utilised in various studies and practices. Six themes and thirty sub-themes emerged from this review.

There are different correlations between health studies and crime studies related to the effects of bias. For example, a systematic literature review found that bias has a positive effect on the patient. However, in crime, bias harms the decision made by the jury or judge, which can cause an innocent criminal to be convicted. Matlock et al (2017) found that cognitive bias was discovered in the decision-making process for ICD implantation. The majority of these biases appeared to favour the use of ICDs. The biases observed in their data not only tended to favour treatment, but it appeared as though many decliners did not fully comprehend what they were declining and that a more informed, less emotional approach to ICD decision making might have swayed some of those patients. On the contrary, (Maroño & Bartels, 2020) argue that participants in the paedophilic condition activated a paedophile schema, causing them to consistently interpret the ambiguous crime scenario schema (i.e., being motivated by the desire to abuse a child sexually). Notably, even if this alternative explanation is correct, it demonstrates how stigmatising beliefs about paedophilia can influence and bias people's judgments.

Most of the effects of bias are focused on teacher assessment (Gweon et al., 2017; Jansen et al., 2021; Marucci et al., 2021; Sheppard, 2019; Yilmaz, 2017) and, to curb this effect, many student task studies suggest using rubrics as one of the practical tools to overcome this problem (Gweon et al., 2017; Jones & Bergin, 2019; Sheppard, 2019; Yilmaz, 2017). Their studies are also more focused on writing tasks (Jansen et al., 2021; Sheppard, 2019), and some of the studies will compare the scores of individual tasks with group tasks to show the effect of bias from raters (Gweon et al., 2017; Yilmaz, 2017).

In the management theme, most studies are more focused on decision-making (Bijani, 2018; Brown et al., 2017; Cantarelli et al., 2020; Matlock et al., 2017; Rego et al., 2018), and some associate it with cognitive bias (Cantarelli et al., 2020; Matlock et al., 2017). They explain that cognitive bias with the Halo effect has a similar concept (Behrmann, 2019; Cantarelli et al., 2020; Marucci et al., 2021; Matlock et al., 2017; Nufer, 2019). Comparisons between genders to look at gender bias (Rego et al., 2018; Sonnert et al., 2018; Xu et al., 2020) is also a subtheme of concern in research as opposed to aged biases (Rego et al., 2018) and attractiveness (Xu et al., 2020). Nonetheless, attractiveness is commonly associated with gender direct and indirectly because the appearance will not be separated from gender. These biases are attributable to the fact that people are frequently distracted by a female's appearance compared to the appearance of a male.

Behavioural patterns primarily determine daily activities. A behaviour pattern leads to a specific action. We have all wondered why some people seem to be so successful. This success is most likely because they have learned to control their impulsive behaviour patterns. Essentially, when determining the presence of bias effects, it is critical to consider behaviour as well. Individuals aware of this bias will be better able to avoid and overcome its consequences (Bijani, 2018; Marucci et al., 2021; Rego et al., 2018; Xu et al., 2020). Those

who have a negative perspective on changes (Bijani, 2018), professional inexperience (Yilmaz, 2017), no direct access, and limited views on group tasks (Gweon et al., 2017), on the other hand, maybe influenced by bias. However, this awareness does not apply in criminal cases, resulting in behaviour that influences decision-making on innocent individuals. As a result, if the judge is unaware of previous individual offences, it is preferable to avoid bias effects (Maroño & Bartels, 2020).

Although rater bias may explain the behaviour, no statistically significant similarities were discovered in the studies examined. This distinction exists because the two are fundamentally dissimilar in this study. There are biases in raters due to untrained raters; however, using numerous raters can result in more reliable analysis (Sonnert et al., 2018). One was because of the apparent clustering of competencies within the same individuals or the illusory Halo and the desired clustering of competencies or the ideal-employee factor (Brown et al., 2017); another was because the rubric was vague, making it difficult to use by inexperienced raters (Bijani, 2018). The other was the rater's inability to understand the criteria, leniency, severe range restriction, and lack of a (Jones & Bergin, 2019).

Training allows employees to acquire new skills, hone existing skills, do better, increase productivity and become better leaders. The cost of training, on the other hand, is relatively high. The need to attend a training programme may cause a project to be delayed. Regardless of potential, training can provide employees and the company with overall benefits that make the time and money invested worthwhile. Training has the potential to improve teaching, cognitive development, and socio-emotional development (Marucci et al., 2021). That is also why training is the most commonly used technique to reduce bias in these studies, but it also depends on one's attitude toward the training. Bias still exists if they initially had a negative attitude toward training (Bijani, 2018).

Future Direction

Other than the Halo and Horn effects, many biases were still unknown and analysed by researchers. In this study, several types of biases were discovered, including reversed Halo effect. Accordingly, several areas of research need to be given attention. First, most current articles in this review are fully quantitative (12). Three studies are entirely qualitative, while the remaining studies (2) have relied on a mixed-method approach. Future studies should consider using more of a mixed-method approach. A mixed-method perspective is a relatively new research paradigm that advocates for the systematic integration, or "mixing," of quantitative and qualitative data within the context of a single investigation or long-term research program. Mixed-methods research helps resolve discrepancies between quantitative and qualitative findings that reflect the perspectives of participants. Mixed-methods research provides a voice for study participants and demonstrates that study findings are grounded in their lived experiences. Because multiple perspectives illuminate the issues under consideration, mixed methods can facilitate scholarly interaction and enrich the research experiences of researchers (Shorten & Smith, 2017).

Most studies rely on electronic keyword searches, which are widely accepted as the optimal method for conducting systematic reviews. However, there are several complementary techniques that researchers can consider when conducting their searches. Citation tracking is one technique that can be considered. It refers to efforts to identify articles related to the ones being studied based on those that cite the papers being studied. This technique enables researchers to follow research leads in both directions. Additionally, the search results can be enriched because they may identify additional publications not

identified via standard database searches due to a search strategy's or bibliographic record's vocabulary constraint (Wright et al., 2014).

Another technique is reference searching, which involves examining the reference list in the selected articles for additional articles. Horsley et al (2011) added that when researchers have difficulty locating pertinent information, examining reference lists may help reduce the risk of missing critical information. Snowballing is another technique that can be classified into two types: forward snowballing and backwards snowballing. One of the primary drawbacks of citation tracking, reference searching, and snowballing systems is that they can spiral out of control, retrieving more articles than are manually appraisable (Tsafnat et al., 2014). Contacting an expert is another technique worth considering, mainly if the specialist literature is unclear (Gotzsche, 2012). In general, researchers may limit their search for bias, Halo effect, and Horn effect to a specific field, such as education, so that the results can be digested and parsed in-depth using the techniques described above, ensuring that no study is overlooked.

Conclusion

The systematic review demonstrated the influence of bias, Halo, and Horn on raters when evaluating students, employees, patients, products, services, teaching and learning, and sports clubs. Each study established these effects through comparison, correlation, outcomes, and primary effect analysis. Additionally, the studies suggested techniques, strategies, processes, and tools for mitigating these bias, Halo, and Horn effects. Based on their systematic reviews, the authors identified six major themes in bias, Halo, and Horn effect: teachers and teaching, students' tasks, management, crime, psychology, and raters. These effects were expanded to include 30 sub-themes. Additionally, the review makes several recommendations for future research. First, future studies should employ a more mixed-method approach, which allows for in-depth analysis of how the effect of bias, Halo, and Horn influenced individual decision-making or judgment and problem-solving to address this issue. Second, complementary techniques such as citation tracking, reference searching, snowballing or contacting experts should be used.

Researchers can use this systematic review to develop their theoretical and conceptual frameworks critically to situate any study about the evaluation of student tasks in order to gain the researchers' process insights, discrepancies, and alternative references by shaping the study's methodologies and design in partnership with the research question, philosophic, paradigmatic, and gaps of study. Researchers can also base the underpinnings of their study on existing research, to draw conclusions, mapping out the variables and the interplay between them. In order to analyse and explain pertinent topics and show how the researcher may incorporate their study into existing research, this paper can establish the essential notion of bias, including two factors, the horn and halo effects.

Moreover, this study can make a significant contribution to knowledge by identifying and resolving issues with bias, halo, and horn effects utilising a different strategy than that of other researchers. The results will definitely add to an existing body of knowledge. For context, this study produced numbers in a vacuum rather than providing the context that will allow for a meaningful analysis of the findings. The statement and ideas are presented in a context that the concept of a bias can fully comprehend. It also can provide the background of a bias study to the informed, as discussed throughout the systematic literature review.

Acknowledgement

We thank our colleagues from Faculty of Language and Communication who provided insight and expertise that greatly assisted the research. Thanks also to anyone I've forgotten who was instrumental in this project

Conflict of Interest Statement

The authors report there are no competing interests to declare

References

- Behrmann, L. (2019). The Halo effect as a teaching tool for fostering research-based learning. *European Journal of Educational Research*, 8(2), 433–441. <https://doi.org/10.12973/eujer.8.2.433>
- Bijani, H. (2018). Investigating the validity of oral assessment rater training program: A mixed-methods study of raters' perceptions and attitudes before and after training. *Cogent Education*, 5(1), 1-20. <https://doi.org/10.1080/2331186X.2018.1460901>
- Brown, A., Inceoglu, I., & Lin, Y. (2017). Preventing Rater Biases in 360-Degree Feedback by Forcing Choice. *Organizational Research Methods*, 20(1), 121–148. <https://doi.org/10.1177/1094428116668036>
- Chantal, Y., Bernache-Assollant, I., & Schiano-Lomoriello, S. (2013). Examining a negative Halo effect to anabolic steroids users through perceived achievement goals, sports personship orientations, and aggressive tendencies. *Scandinavian Journal of Psychology*, 54(2), 173–177.
- Chartered Institute of Personnel and Development (CIPD). (2015). *Championing better work and working [Annual Review]* p.p 20-28. https://www.cipd.co.uk/Images/annual-review_2014-15_tcm18-11771.pdf
- Cantarelli, P., Belle, N., & Belardinelli, P. (2020). Behavioral Public HR: Experimental Evidence on Cognitive Biases and Debiasing Interventions. *Review of Public Personnel Administration*, 40(1), 56–81. <https://doi.org/10.1177/0734371X18778090>
- Dompnier, B., Pansu, P., & Bressoux, P. (2006). An integrative model of scholastic judgments: Pupils' characteristics, class context, Halo effect and internal attributions. *European Journal of Psychology of Education*, 21(2), 119-133. doi:10.1007/BF03173572
- Ehrlinger, J., Gilovich, T., & Ross, L. (2005), "Peering into the bias blind spot: people's assessments of bias in themselves and others", *Personality and Social Psychology Bulletin*, 31 (5), 680-692.
- Forgas, J. P., & Laham, S. M. (2017). Halo effects. In R. F. Pohl (Ed.), *Cognitive illusions: Intriguing phenomena in judgement, thinking and memory* (2nd ed., pp. 276–290). New York, NY: Psychology Press
- Gest, F. S. D. (2006). Teacher reports of children's friendships and social groups: Agreement with peer reports and implications for studying peer similarity. *Social Development*, 15, 248-259. doi:10.1046/j.1467-9507.2006.00339.x
- Gibson, J. L., & Gore, J. S. (2015). You're OK until you misbehave: How norm violations magnify the attractiveness devil effect. *Gender Issues*, 32(4), 266–278. <https://doi.org/10.1007/s12147-015-9142-5>
- Graf, M., & Unkelbach, C. (2016). Halo effects in trait assessment depend on information valence: Why being honest makes you industrious, but lying does not make you lazy. *Pers Soc Psychol Bull.* 42(3):290-310. doi: 10.1177/0146167215627137. Epub 2016 Jan 25. PMID: 26811437.

- Gotzsche, P. C. (2012). Content area experts as authors: helpful or harmful for systematic reviews and meta-analyses? *BMJ* 2012 (345), e7031
- Gweon, G., Jun, S., Finger, S., & Rose, C. P. (2017). Towards effective group work assessment: even what you don't see can bias you. *International Journal of Technology and Design Education*, 27(1), 165–180. <https://doi.org/10.1007/s10798-015-9332-1>
- Horsley, T., Dingwall, O., & Sampson, M. (2011). *Examining reference lists to find relevant studies for systematic reviews*. http://www.cochrane.org/MR000026/METHOD_examining-reference-lists-to-find-relevant-studies-for-systematic-reviews.
- Jackson, D. A., & King, A. R. (2004). Gender differences in the effects of oppositional behavior on teacher ratings of ADHD symptoms. *Journal of Abnormal Child Psychology*, 32(2), 215–224. <https://doi.org/10.1023/B:JACP.0000019772.71251.ff>
- Jansen, T., Vogelin, C., Machts, N., Keller, S., & Moller, J. (2021). Don't just judge the spelling! the influence of spelling on assessing second-language student essays. *Frontline Learning Research*, 9(1), 44–65. <https://doi.org/10.14786/flr.v9i1.541>
- Jones, E., & Bergin, C. (2019). Evaluating Teacher Effectiveness Using Classroom Observations: A Rasch Analysis of the Rater Effects of Principals. *Educational Assessment*, 24(2), 91–118. <https://doi.org/10.1080/10627197.2018.1564272>
- Kahneman, J. D. (2011). *Thinking, fast and slow*. London: Allen Lane.
- Landy, D., & Sigall, H. (1974). Beauty is talent: Task evaluation as a function of the performer's physical attractiveness. *Journal of Personality and Social Psychology*, 29(3), 299–304. <https://psycnet.apa.org/record/1974-25236-001>
- Marono, A., & Bartels, R. M. (2020). Examining the judgments of pedophiles in relation to a non-sexual offense. *Psychology, Crime and Law*, 26(9), 887–901. <https://doi.org/10.1080/1068316X.2020.1742339>
- Marucci, E., Oldenburg, B., Barrera, D., Cillessen, A. H. N., Hendrickx, M., & Veenstra, R. (2021). Halo and association effects: Cognitive biases in teacher attunement to peer-nominated bullies, victims, and prosocial students. *Social Development*, 30(1), 187–204. <https://doi.org/10.1111/sode.12455>
- Matlock, D. D., Jones, J., Nowels, C. T., Jenkins, A., Allen, L. A., & Kutner, J. S. (2017). Evidence of Cognitive Bias in Decision Making Around Implantable-Cardioverter Defibrillators: A Qualitative Framework Analysis. *Journal of Cardiac Failure*, 23(11), 794–799. <https://doi.org/10.1016/j.cardfail.2017.03.008>
- Myford, C. M., & Wolfe, E. W. (2004). Detecting and measuring rater effects using many-facet Rasch measurement: Part II. *Journal of Applied Measurement*, 5(2), 189–227.
- Nisbett, R. E., & Wilson, T. D. (1977). The Halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology*, 35, 250–256. https://www.researchgate.net/publication/232501672_The_Halo_Effect_Evidence_for_Unconscious_Alteration_of_Judgments
- Nufer, G. (2019). "Say hello to Halo": The Halo effect in sports. *Innovative Marketing*, 15(3), 116–129. [https://doi.org/10.21511/im.15\(3\).2019.09](https://doi.org/10.21511/im.15(3).2019.09)
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hrobjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., Whiting, P., & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting

- systematic reviews. *Systematic Reviews*, 10 (89), 1-11. <https://doi.org/10.1186/s13643-021-01626-4>
- Rego, A., Vitoria, A., Tupinamba, A., Junior, D. R., Reis, D., Cunha, M. P., & Lourenço-Gil, R. (2018). Brazilian managers' ageism: a multiplex perspective. *International Journal of Manpower*, 39(3), 414–433. <https://doi.org/10.1108/IJM-09-2016-0167>
- Roelofs, S., Boleij, H., Nordquist, R. E., & Van der Staay, F. J. (2016). Making decisions under ambiguity: Judgment bias tasks for assessing emotional state in animals. *Frontiers in Behavioral Neuroscience*, 10 (119), 1-16. https://www.researchgate.net/publication/303867370_Making_Decisions_under_Ambiguity_Judgment_Bias_Tasks_for_Assessing_Emotional_State_in_Animals
- Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (pp.173–220). New York: Academic Press.
- Sheppard, B. (2019). Checking Your Analytic Performance Rubrics for a Halo Effect. *ORTESOL Journal*, 36, 28–33. ERIC - EJ1219915 - Checking Your Analytic Performance Rubrics for a Halo Effect, ORTESOL Journal, 2019 (ed.gov)
- Shifrer, D. (2013). Stigma of a label: Educational expectations for high school students labeled with learning disabilities. *Journal of Health and Social Behavior*, 54(4), 462–480. <https://journals.sagepub.com/doi/full/10.1177/0022146513503346>
- Shorten, A., & Smith, J. (2017). Mixed methods research: expanding the evidence base. *Evidence-Based Nursing* 20, 74-75.
- Sonnert, G., Hazari, Z., & Sadler, P. M. (2018). Evaluating the quality of middle school mathematics teachers, using videos rated by college students. *Studies in Educational Evaluation*, 58(May), 60–69. <https://doi.org/10.1016/j.stueduc.2018.05.006>
- Steglich, C., & Knecht, A. (2014). Studious by association? Effects of teachers' attunement to students' peer relations. *Zeitschrift für Erziehungswissenschaft*, 17(S5), 153-170. doi:10.1007/s11618-014-0556-
- Sutu, A., Serrano, S., Schultz, L. H., Jackson, J. J., & Damian, R. I. (2019). Creating Through Deviancy or Adjustment? The Link between Personality Profile Normativeness and Creativity. *European Journal of Personality*, 33(5), 565–588. <https://doi.org/10.1002/per.2215>
- Tariq, M., Govaerts, M., Afzal, A. Ali, S. A., & Zehra, T. (2020). Ratings of performance in multisource feedback: comparing performance theories of residents and nurses. *BMC Med Educ.* 20, 355 <https://doi.org/10.1186/s12909-020-02276-1>
- Thorndike, E. L. (1920). A constant error in psychological ratings. *Journal of Applied Psychology*, 4, 25-29. (PDF) A constant error in psychological ratings. (1920) | E.L. Thorndike | 1572 Citations (typeset.io)
- Tsafnat, G., Glasziou, P., Choong, M. K., Dunn, A., Galgani, F., & Coiera, E. (2014). Systematic review automation technology. *Systematic Review* 9 (3), 74. Systematic review automation technologies - PubMed (nih.gov)
- Wright, K., Golder, S., & Rodriguez-Lopez, R. (2014). Citation searching: a systematic review case study of multiple risk behaviour interventions. *BMC Medical Research Methodology* 14(73), 1–8. <https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-14-73>
- Xu, S., Martinez, L., & Smith, N. A. (2020). The effects of attractiveness, gender and self-esteem in service jobs. *International Journal of Contemporary Hospitality Management*, 32(1), 249–266. <https://doi.org/10.1108/IJCHM-02-2019-0127>

Yilmaz, F. (2017). Analysis of the rater effects in the rating of diagnostic trees prepared by teacher candidates by the Many-Facet Rasch Model. *Journal of Education and Practice*, 8(18), 174–184. <https://www.researchgate.net/publication/333162837>