DOI: 10.47743/ejpar.2026-6-6

INSIDE THE BLACK BOX: MEASURING THE PERFORMANCE-RECOGNITION GAP AMONG WOMEN IN ISRAELI HIGH-TECH

ZEALA PINTO
Doctoral School of Economics and Business Administration
Alexandru Ioan Cuza University of Iasi
Iași, Romania
tsaela@gmail.com

Abstract
Women enter the Israeli high-tech pipeline in substantial numbers yet remain sharply underrepresented in senior technical and leadership roles, holding fewer than 12% of chief technology officer positions in the sector's high-growth domains (Israel Innovation Authority, 2025). This paper examines the Performance-Recognition Gap, defined as the systematic tendency for women's professional contributions to receive less credibility, visibility, and reward than equivalent work by male colleagues. Drawing on Heilman's (2001) lack-of-fit model, Castilla and Benard's (2010) meritocracy paradox, and the double bind framework (Eagly and Karau, 2002; Rudman and Glick, 2001), the study draws on qualitative findings from exploratory interviews that informed an adapted survey-based operationalization tested among 220 women in Israeli high-tech. The Performance-Recognition Gap construct, adapted from the Gender Bias Scale for Women Leaders (Diehl et al., 2020), showed high internal consistency (Cronbach's alpha = 0.866) and is interpreted through three related facets: the Credibility Tax, Transparent Effort, and the Broken Scale. Rather than treating recognition asymmetry as a single undifferentiated experience, the paper approaches it as an evaluative mechanism and examines its internal structure. Quantitative findings provide triangulating support: the construct is associated with lower relative career advancement, lower advancement expectancy, and higher exit intention. This article presents an initial adapted operationalization rather than a full psychometric validation.
Keywords: Performance-Recognition Gap; gendered evaluation; Israeli high-tech; meritocracy paradox; double bind.
JEL Classification: J16; J24; J71; M12; O33.

1. INTRODUCTION

This paper examines the Performance-Recognition Gap as an evaluative mechanism operating through everyday organizational practices. Rather than treating recognition asymmetry only as an outcome pattern, the paper approaches it from within, through three related facets: the credibility accorded to women's contributions (the Credibility Tax), the visibility of relational and facilitative effort (Transparent Effort), and the translation of assessed

performance into reward (the Broken Scale). By unpacking these dimensions, the paper aims to open the black box of how evaluation systems produce gendered recognition outcomes.

Women in technology are promoted more slowly (McKinsey and Company, 2024), paid less (PayScale, 2024), and exit the sector at higher rates than their male peers (AnitaB.org, 2022). The mechanisms through which these outcomes are produced, however, remain less well measured. Global data illustrate the scale of the structural gap. For every 100 men promoted to their first management role, only 87 women receive the same advancement, a disparity McKinsey and Company (2024) term the "broken rung." The World Economic Forum (2022) estimates that, at current rates of progress, closing the global economic gender gap will require more than a century. In Israel, these patterns are reproduced within one of the world's most concentrated and internationally scrutinised technology ecosystems: women represent 28% of research and development employees but hold fewer than 12% of chief technology officer positions in high-growth domains (Israel Innovation Authority, 2025). Women in the sector also receive only 4.3% of venture capital funding (Israel Innovation Authority, 2025), reflecting barriers to capital access alongside the cumulative effects of informal professional networks that shape sponsorship, visibility, and advancement.

The standard organizational response to these disparities has tended to focus on downstream interventions, including mentoring programs, flexible work arrangements, and targeted recruitment at pipeline entry points. While these measures may address retention pressures, they leave upstream evaluative processes less directly examined. This paper examines one such process: the Performance-Recognition Gap, defined as the systematic tendency for women's professional contributions to receive less credibility, visibility, and reward than equivalent work by male colleagues through everyday evaluation practices that appear objective but are shaped by gendered standards of assessment. The Performance-Recognition Gap refers to the evaluative process through which salary and seniority disparities may be reproduced, rather than to those outcome patterns directly.

The Israeli high-tech context sharpens this analysis two ways. First, the sector often operates through compressed evaluation cycles in which technical contributions are assessed rapidly, informally, and continuously, amplifying the influence of implicit standards and social expectations on who receives credit for what. Second, the sector's founding mythology is strongly meritocratic. The belief that "the best code wins" functions as a cultural article of faith. Research on the meritocracy paradox shows that organizations with the most explicit meritocratic rhetoric can exhibit greater gender bias in reward allocation than those making no such claims, partly because belief in a fair system reduces the impulse to check for unfairness (Castilla and Benard, 2010). In an ecosystem where meritocracy is both celebrated and monetised, the Performance-Recognition Gap may function as a hidden tax on women's contributions.

2. THEORETICAL BACKGROUND

2.1. Gendered Evaluation Systems in High-Tech Leadership

Women's underrepresentation in technology leadership has often been attributed to pipeline entry challenges, including lower participation in technical degree programs and lower recruitment into technical roles. While these challenges are well documented, this paper focuses on a different point in the career trajectory: the evaluation and recognition systems inside organizations that shape whose contributions are credited, whose ideas advance, and whose output converts into pay and promotion. Global data illustrate the scale of the resulting gap. McKinsey and Company (2024) report that for every 100 men promoted to the first management level, only 87 women receive equivalent advancement. This "broken rung" is consequential because it creates a cumulative numerical deficit at the base of the leadership structure that is difficult to correct through downstream diversity programs alone. In Israel, the same pattern is visible within one of the world's most concentrated and internationally scrutinised technology ecosystems: women constitute 28% of research and development employees across the technology sector, yet hold fewer than 12% of chief technology officer positions in high-growth domains and receive only 4.3% of venture capital funding (Israel Innovation Authority, 2025). Research consistently places some of the steepest attrition at early promotion points, where advancement is shaped by formal criteria alongside informal assessments of technical credibility and leadership readiness (McKinsey and Company, 2021; McKinsey and Company, 2024).

Controlled experimental designs show that evaluation systems can introduce gender bias before any formal decision is made. Steinpreis, Anders, and Ritzke (1999) presented identical curriculum vitae to faculty evaluators, varying only the name assigned to the applicant. The same document, when attributed to a male applicant, was rated as significantly more qualified for a tenure-track academic position than when attributed to a female applicant. Both male and female faculty showed the same pattern, pointing to shared cultural schemas that assign different baseline assumptions of competence and authority to male and female candidates. Uhlmann and Cohen (2005) demonstrated a parallel phenomenon in organizational hiring decisions: evaluators did not apply a fixed definition of merit, but shifted the criteria used to justify a decision after the candidate had been identified. When the preferred candidate was male, experiential criteria were weighted more heavily; when equivalent qualifications were assigned to a female candidate, the same evaluators privileged formal credentials. Merit was thus constructed around the preferred candidate rather than applied neutrally across candidates. This shifting of criteria is especially relevant to the Performance-Recognition Gap because it helps render bias less

visible to evaluators themselves: each individual decision may appear principled even when the aggregate pattern systematically disadvantages women.

Field evidence suggests that similar biases can operate at the rewards stage even when the evaluation stage appears to function equitably. Castilla (2008) analyzed longitudinal compensation data from a large service organization and found that women and racial minorities who received identical performance ratings to their white male colleagues nonetheless received smaller salary increases. The conversion of ratings into rewards was itself gendered. Managers exercised discretionary judgment at the compensation stage in ways that eroded the parity established at the evaluation stage. Castilla's conclusion is direct: meritocratic evaluation procedures do not necessarily produce meritocratic reward allocation, and the gap between the two is a key site for examining performance-recognition dynamics. This logic informs the salary-comparison items used in the present study, which capture perceived divergence between assessed performance and compensation.

2.2. Theoretical Mechanisms: Lack-of-Fit, the Meritocracy Paradox, Bias in Ambiguous Contexts, and the Double Bind

Lack-of-Fit: Stereotypes as a Pre-Performance Evaluation Filter

Heilman's (2001) lack-of-fit model provides the foundational framework for understanding how gender stereotypes systematically distort performance evaluation in technical and leadership roles. The model identifies two distinct processes, both of which disadvantage women, but through different routes. Descriptive stereotypes establish a perceived mismatch between the attributes typically associated with women (communality, warmth, interpersonal deference) and the attributes considered necessary for success in technical and leadership positions (analytic rigor, decisiveness, independent authority). This perceived lack of congruence leads evaluators to expect women to perform less well, and to filter incoming performance evidence through that prior expectation. A woman's achievement that would confirm competence in a male colleague is interpreted as an exception, an overperformance, or a team effect, rather than as reliable evidence of capability. As Heilman articulated, "gender stereotypes not only describe how men and women actually differ... they also prescribe how men and women should differ" (Heilman, 2001, p. 658). This prescriptive dimension generates the second process: when women do succeed in male-typed domains, they violate the norm of femininity and face what Heilman terms backlash, a penalty imposed not for poor performance but for gender-norm violation. The dual architecture of the lack-of-fit model explains why conforming to feminine expectations reduces competence perceptions, while demonstrating technical competence risks social penalty. Both pathways reduce the probability that women's contributions will be fully credited. The items in

the present study capturing unequal standards and heightened scrutiny map directly onto this mechanism: they measure the experienced daily effect of evaluation practices shaped by default lack-of-fit assumptions.

The Meritocracy Paradox: Fairness Rhetoric as a Bias Amplifier

Castilla and Benard (2010) extended this analysis by demonstrating, through experimental manipulation of organizational contexts, that the explicit emphasis on meritocracy as an institutional value paradoxically increases gender bias in reward allocation. In organizations whose systems were explicitly labeled as merit-based, managers awarded larger performance bonuses to male employees than to equally performing female employees. The effect was absent or reduced in control conditions where meritocratic values were not primed. The mechanism Castilla and Benard identify is moral credentialing. When managers operate within systems they believe to be fair, they experience a reduced motivation to monitor their own evaluation behavior for bias. The institutional belief in a meritocratic process functions as a moral license to act on implicit preferences without recognizing those preferences as bias. This finding carries substantial weight in the Israeli high-tech context, where meritocratic ideology is deeply embedded in organizational identity. The cultural premise that code quality and technical output are objective and gender-neutral creates precisely the conditions under which moral credentialing operates: evaluators are least likely to audit their own practices for bias when they are most certain that no bias exists. The moral credentialing effect operates with particular force on one category of contribution documented in this study: the facilitative, harmony-maintaining work that women disproportionately perform in team settings. Because evaluation systems oriented toward individually attributed technical outputs have no accounting category for relational labor, this work passes through the system without generating a credit entry regardless of its value to collective performance. The contribution is real and visible to those who receive it; the accounting system simply has no mechanism for registering it. Salary outcomes relative to male peers provide the downstream marker where these unregistered contributions ultimately accumulate.

Bias in Ambiguous Evaluation Contexts: Conscious and Unconscious Forms

A substantial body of research has established that bias in evaluation does not require deliberate intent. Dovidio and Gaertner (2000) demonstrated that evaluative discrimination concentrates in the ambiguous middle of assessment: when candidate qualifications are clearly strong or clearly weak, explicit fairness norms constrain evaluator behavior, but when credentials are ambiguous or mixed, implicit preferences fill the interpretive gap. Critically, this form of bias is most pronounced among evaluators who explicitly hold egalitarian values and who believe themselves to be unbiased. The evaluator's self-perception of

fairness, rather than eliminating bias, reduces the vigilance that might otherwise constrain it. In high-technology evaluation contexts, where technical contributions routinely require contextual judgment about the difficulty of a problem, the originality of a solution, or the quality of an architectural decision, that ambiguity is the norm rather than the exception. Every judgment call about the relative merit of an idea, a design choice, or a leadership contribution is an ambiguous evaluation moment. Dovidio and Gaertner's (2000) framework explains why women report that their ideas gain traction when repeated by a male colleague, and that their performance is held to a higher evidentiary standard: in each case, the gap is not a product of calculated exclusion but of implicit criteria applied unevenly in moments where discretion is unavoidable. The implication is that reducing the Performance-Recognition Gap requires structural changes to how evaluation decisions are made, not merely awareness training aimed at individual intentions.

The Double Bind: A Structural Trap With No Exit

Eagly and Karau (2002) formalized the logical endpoint of the mechanisms described above as role congruity theory. Their account holds that perceived incongruity between the communal attributes ascribed to women and the agentic attributes associated with leadership roles produces two forms of evaluative prejudice. The first reduces the positive evaluation of women as potential leaders before any leadership behavior has been observed. The second penalizes leadership behavior itself when enacted by women, because competent assertion in women violates the communal norm and triggers what Rudman and Glick (2001) term backlash: negative social and professional consequences for women who display the very behaviors that would be rewarded in a male colleague. The resulting structure is a trap from which there is no logical exit. Women who conform to the communal gender role satisfy social expectations but are perceived as insufficiently competent for technical or leadership authority. Women who perform the agentic behaviors required for recognition and advancement violate the communal norm and face interpersonal and reputational penalties. In both cases, the performance-recognition mechanism operates: contribution is either discounted because the woman is seen as the wrong type, or penalized because she is seen as the wrong kind of woman. Eagly and Karau (2002) state the problem directly: "Women face a double bind: if they conform to their gender role, they are seen as weak leaders; if they conform to the leader role, they are seen as unfeminine and penalized for their behavior" (p. 576). The double bind provides the structural architecture that connects each of the three facets of the Performance-Recognition Gap: unequal credibility standards, invisible relational contributions, and salary gaps that survive equal performance all follow from the same no-win evaluation logic. Section 5.2 discusses this interpretation further.

2.3. Measuring the Performance-Recognition Gap: Scale and Instrument

The measurement instrument used in this study is an adaptation of the Gender Bias Scale for Women Leaders (GBSWL), developed and validated by Diehl, Stephenson, and Dzubinski (2020) across multiple industries in the United States. The GBSWL is a 47-item instrument organized around six higher-order categories of gender bias experienced by women in professional leadership contexts. Its development followed systematic qualitative and quantitative validation procedures across independent samples, producing subscale reliability coefficients above 0.70 across most factors. The present study uses the GBSWL as its measurement base, applied to the Israeli high-tech context.

The Performance-Recognition Gap construct comprises items from the adapted instrument capturing experienced disparities along three empirically distinguishable dimensions. The first concerns credibility and evaluation standards, the second concerns the visibility of collaborative labor and relational effort, and the third concerns the translation of performance into formal rewards and advancement outcomes. These conceptual, contextual, and measurement anchors are summarised (table 1).

Table 1. Conceptual, Contextual, and Measurement Anchors for the Performance-Recognition Gap

Anchor	Year	Core mechanism	Evidence type
Heilman, M.E.	2001	Lack-of-fit: descriptive and prescriptive stereotypes distort evaluation before performance is observed	Theoretical, experimental
Castilla, E.J.	2008	Identical performance ratings produce unequal salary outcomes: reward translation is gendered	Longitudinal field study
Castilla, E.J. and Benard, S.	2010	Meritocracy paradox: explicit fairness rhetoric reduces evaluator self-monitoring and increases bias	Experimental
Steinpreis, R.E. et al.	1999	Identical credentials rated lower under female name by both male and female evaluators	Experimental (CV paradigm)
Uhlmann, E.L. and Cohen, G.L.	2005	Criteria shift post-hoc to favor male candidates; merit is constructed not applied	Experimental
Dovidio, J.F. and Gaertner, S.L.	2000	Implicit bias concentrates in ambiguous evaluation; strongest among self-perceived fair evaluators	Experimental
Eagly, A.H. and Karau, S.J.	2002	Role congruity theory: communal-agentic incongruity produces double bind with no exit	Theoretical, meta-analytic
Rudman, L.A. and Glick, P.	2001	Backlash against agentic women: prescriptive stereotype violation triggers social and professional penalties	Experimental

Anchor	Year	Core mechanism	Evidence type
McKinsey and Company	2024	Broken rung: 87 women promoted per 100 men at first management transition	Industry report
Israel Innovation Authority	2025	28% R&D, under 12% CTOs, 4.3% VC: representation declines sharply at each leadership transition	Government report
Diehl, A.B. et al.	2020	Gender Bias Scale for Women Leaders (GBSWL): validated 47-item instrument across US industries	Psychometric validation

Source: authors compilation

3. METHOD

3.1. Research Design

This study uses a cross-sectional quantitative survey design. The survey was informed by exploratory semi-structured interviews (n = 19) with women in the sector, conducted prior to survey administration to contextualize how evaluation and recognition experiences are perceived in the Israeli high-tech environment. Following survey data collection, two post-survey focus groups were conducted with women from the same population to contextualize the quantitative patterns and enable collective reflection on the survey findings. The present paper reports the quantitative findings, drawing on the interview and focus-group material to illustrate and situate the survey results.

3.2. Survey Instrument and Sample

The quantitative instrument was adapted from the Gender Bias Scale for Women Leaders (GBSWL), developed by Diehl, Stephenson, and Dzubinski (2020) across multiple industries in the United States. All items were translated and contextually tailored for the Israeli technology sector, and rated on a five-point Likert scale ranging from 1 (Strongly Disagree) to 5 (Strongly Agree).

The Performance-Recognition Gap was examined using selected items from the adapted instrument reflecting three related areas: the credibility accorded to women's contributions, the visibility of their collaborative and relational labor, and the translation of performance into promotion and reward outcomes. Reliability was strong across the three facets and the full nine-item construct: Credibility Tax (k = 4, ? = .827), Transparent Effort (k = 3, ? = .793), Broken Scale (k = 2, ? = .860), and the full construct (? = .866). Full item wordings are provided in the Appendix.

The analyses in this paper represent an initial adapted operationalization of the construct for the Israeli high-tech context rather than a full psychometric validation study. Confirmatory factor analysis and convergent validity testing are identified in Section 6.2 as directions for future research. A total of 224 responses were submitted; after data cleaning, the final sample comprised 220 women. Participants were recruited through professional networks, LinkedIn

outreach, and snowball sampling. The survey was administered online, and participation was entirely voluntary. Eligibility required that respondents identify as women, be aged 18 or over, and be currently employed in or recently departed from Israel's high-tech sector. The sample covers three functional clusters: core technical roles including research and development, engineering, and data science (n = 131); product and design roles (n = 50); and strategic support functions within high-tech organizations including sales, marketing, and corporate operations (n = 39). The sample spans the full organizational hierarchy from individual contributor to C-level executive. The median age was 41 and the average tenure was 4.9 years. Outcome variables, including relative career advancement, advancement expectancy, professional safety perception, and exit intention, were each measured using single-item five-point scales.

3.3. Analytical Strategy

The quantitative analysis was structured to address the three research questions in sequence. First, the study examined whether the Performance-Recognition Gap was present and reliably measurable in this population through descriptive and internal consistency analysis (Section 4.1). Second, it examined the internal facets of the construct as an interpretive framework for understanding recognition asymmetry (Section 4.2). Third, it considered whether the resulting pattern was broadly consistent with selected career outcomes (Section 4.3).

3.4. Ethical Considerations

All participants provided informed consent before completing the survey. Data were anonymized at the point of collection: no identifying information was retained in the analytical dataset. Participation was entirely voluntary, and respondents were informed of their right to withdraw at any time before submission without consequence. The anonymous survey format was selected to reduce social desirability effects given the sensitivity of reporting workplace bias experiences.

4. FINDINGS

This section presents findings in three stages: first, empirical presence and coherence; second, interpretive structure; third, quantitative associations.

4.1. A Coherent Pattern of Recognition Asymmetry

The Performance-Recognition Gap construct demonstrated strong internal consistency in the present sample (Cronbach's alpha = 0.866), meeting the threshold established by Hair et al. (2009). The construct mean of 2.728 on the five-point scale, with 32.2% of respondents endorsing items at the top-box level (scores of 4 or 5), indicates a moderate level of perceived evaluative credit disparity. The mean falls below the scale midpoint of 3, reflecting that

endorsement of these experiences is not universal across the sample; nonetheless it sits substantially above the lower bound of the scale, and no item mean drops below 2.3, indicating that the experiences captured are distributed across the sample rather than confined to a small minority of respondents. Across items, the pattern is not uniform: items addressing salary outcomes and effort expectations received higher mean endorsement than items addressing contribution visibility and relational work. This distribution indicates that the gap operates across more than one moment in the evaluation cycle, spanning the credibility attached to women's individual contributions, the recognition of their collaborative work, and the translation of assessed performance into compensation, rather than expressing a single generalised perception of unfairness.

This statistical coherence is consistent with exploratory interviews conducted before the survey. Participants described recognition asymmetry as a recurring organizational pattern embedded in everyday evaluation moments, including calibration meetings, promotion discussions, and informal credit attribution in team settings.

Facet-level internal consistency analysis revealed variation across the three dimensions. The credibility-standards dimension demonstrated strong cohesion (Cronbach's alpha = .827), as did the compensation-comparison dimension (Cronbach's alpha = .860). For this two-item facet, the inter-item correlation was r = .754, further supporting its internal coherence. The effort-visibility dimension required closer examination. Initial reliability was Cronbach's alpha = .701, which, while above the minimum acceptable threshold, fell below the preferred level. Item-level diagnostics identified one item, a reverse-coded statement about whether relational harmony contributions receive recognition. Its corrected item-total correlation was .19, below the .30 threshold specified by Hair et al. (2009). Removing this item raised facet reliability to Cronbach's alpha = .793. All subsequent analyses of the effort-visibility dimension use this three-item version.

4.2. Opening the Black Box: Three Facets of Recognition Asymmetry

The item-level pattern reported above establishes that the Performance-Recognition Gap operates across three related dimensions. Their interpretation is the focus of this section. The facet labels proposed below are interpretive summaries of closely related item clusters that emerged from reliability analysis: they are offered as a conceptual framework for organizing the item-level evidence rather than as claims about fully distinct latent dimensions, which would require confirmatory factor analysis to establish.

The first facet, the Credibility Tax (M = 2.84, top-box = 36.4%, α = .827, k = 4), encompasses items capturing the surplus evidence of competence that women must supply to earn the same professional credibility as equivalent male colleagues. This cluster describes heightened scrutiny, a higher effort threshold for recognition, an expectation of communal rather than agentic conduct, and the

pattern whereby women's ideas obtain organizational traction primarily after being reiterated by a male colleague. The Credibility Tax is the most consistently reported facet. Interview participants described needing to supply substantially more evidence than male peers to obtain equivalent acceptance, a persistent cognitive load as women calibrated their behavior to avoid penalties for authority displays that would be unremarkable in male colleagues, and promotion calibration processes that evaluated female candidates on demonstrated assertiveness rather than delivered outcomes. These accounts indicate that the credibility surplus cost captured in the survey items is a recurring feature of evaluation practice.

The second facet, Transparent Effort (M = 2.61, top-box = 24.7%, α = .793, k = 3), captures the relational and collaborative work that women disproportionately perform but that formal evaluation systems fail to register as a professional contribution. The work is visible in its effects (teams function, conflicts are absorbed, junior colleagues are supported), but it passes through the evaluation accounting system without generating a credit entry. Items in this cluster describe interrupted speech, suppressed voice in majority-male settings, and unrecognized collegial assistance. Interview participants described evaluation systems that rewarded physical presence over output efficiency, described absorption into non-promotable facilitative labor while male colleagues focused on career-advancing work, and noted that informal decision-making excluded those outside dominant social networks ("the key decisions happened over lunch breaks, where I wasn't present"). Focus-group participants indicated that facilitative contributions, including meeting coordination, onboarding of new team members, and conflict resolution, were expected of women yet systematically omitted from performance reviews, reflecting a structural property of evaluation systems designed around individually attributed technical outputs.

The third facet, the Broken Scale (M = 3.09, top-box = 35.2%, α = .860, k = 2), measures the divergence between women's assessed performance and their actual compensation, relative to both current male peers and men who previously held the same role. The salary-comparison items showed the highest endorsement of any items in the construct: the item capturing pay disparity relative to current male colleagues returned the highest mean in the entire construct (M = 3.26), while pay disparity relative to men who previously held the same role returned M = 2.93. The Broken Scale is the facet that crosses above the scale midpoint, indicating that more than half the sample endorsed these experiences. Focus-group and interview accounts converged on a pattern of evaluative asymmetry in formal review processes: women who insisted on quality execution were described as "aggressive" while equivalent behavior in male colleagues was framed as assertiveness and ambition; men were advanced on perceived potential while women were required to demonstrate an established track record; and performance review language encoded communal terms for

women ("collaborative," "nurturing") against agentic terms for men ("strategic thinking") for identical contribution types. Taken together, the qualitative accounts indicate that the Broken Scale captures an evaluative asymmetry in which identical professional behavior is assessed through different standards depending on gender.

The item removed from the effort-visibility dimension during psychometric diagnostics (corrected item-total r = .19) may reflect conditional applicability: the item presupposed active interruption behavior and may therefore capture situational dynamics rather than stable perceptions of relational effort. A consolidated cross-stage summary is presented (Table 2).

Table 2. Multi-Stage Findings Summary: Convergence Across Data Sources

Data source	Credibility Tax	Transparent Effort	Broken Scale
Exploratory interviews (n = 19)	Calibration meetings applied assertiveness criteria to women but not to equivalent male candidates; participants described needing substantially more evidence for equal acceptance	Output quality matched or exceeded peers, yet evaluation weight fell on physical presence and late-hour availability rather than deliverable quality	Promotion decisions favored men's perceived potential over women's demonstrated track record
Survey (n = 220)	Highest facet consistency (α = .827); scrutiny and idea-attribution items endorsed across seniority levels	Facet α = .793 after item removal; relational contributions routinely unregistered in formal reviews	Highest item-level endorsement in construct; salary-comparison items crossed scale midpoint (α = .860)
Focus groups (n = 2 groups)	Professional authority displays labeled "aggressive" in formal reviews; identical behavior in male colleagues framed as ambition	Facilitative work (coordination, onboarding, conflict resolution) expected of women yet omitted from performance reviews	Men advanced on potential, women required established track record; evaluative yardstick differed by gender
Integrated interpretation	Surplus evidence cost documented across all three data sources; mechanism operates through everyday evaluation moments, not isolated incidents	Invisible-labor pattern consistent across interviews, survey, and focus groups; structural feature of individual-productivity evaluation logic	Compensation and promotion asymmetry reflects cumulative downstream consequence of credibility and visibility gaps

4.3. Quantitative Triangulation

The qualitative accounts reported above establish the interpretive structure of the three facets. The quantitative associations reported below provide a complementary form of evidence: they indicate that the construct, as measured by the survey instrument, is broadly consistent with career outcomes in the directions the mechanism account would suggest.

The Performance-Recognition Gap showed significant negative associations with relative career advancement (r = -.494, p < .001) and advancement expectancy (r = -.530, p < .001), and a significant positive association with exit intention (r = .412, p < .001). The association with professional safety perception was also negative and significant (r = -.387, p < .001). The strength and consistency of this pattern across four outcome variables indicates that the construct captures a recognition pattern meaningfully associated with career outcomes rather than a localised perception confined to one domain. The correlation with advancement expectancy was the largest in the dataset, modestly exceeding the correlation with actual career advancement. This ordering is consistent with the interpretation in Section 5.4: an early associated outcome of perceived recognition failures may be a progressively eroded belief that sustained effort will translate into commensurate progress, rather than a stalled career record.

After adjustment for seniority, tenure, and demographic factors, the Performance-Recognition Gap retained a significant negative association with relative career advancement (β = -.190, p = .033). Women at higher organizational levels reported better relative career advancement, consistent with a survival effect in which those who reach senior positions have already cleared the evaluation barriers contributing to attrition at earlier stages.

5. DISCUSSION

5.1. Opening the Black Box: Convergence and Coherence

These findings converge across three distinct data sources (exploratory interviews with nineteen women in Israeli high-tech, a survey of 220 women drawn from the same population, and two focus groups conducted after the survey), producing the same pattern in different registers. Women describe a professional environment in which their contributions meet evaluative standards that shift in ways that disadvantage them, in which the work most likely to be rendered invisible is the work they disproportionately perform, and in which the gap between performance assessed and performance rewarded is a routine feature of career experience rather than a memorable exception. The Performance-Recognition Gap construct, with a Cronbach's alpha of .866, suggests that this pattern forms a coherent evaluative mechanism with a consistent internal structure.

That structure is a central finding of this study. The mechanism operates through three related moments in the evaluation cycle. The Credibility Tax operates at the point where contribution is first assessed: the threshold of evidence required before a woman's work is credited as competent. Transparent Effort operates at the accounting stage, where the evaluation system decides which categories of work generate a professional credit entry and which pass through without one. The Broken Scale operates at the translation stage, where assessed performance is converted into compensation and advancement. With 32.2% of respondents endorsing items at the top-box level and facet means ranging from 2.61 (Transparent Effort) to 3.09 (Broken Scale, the only facet to cross the scale midpoint), the data suggest that the mechanism distributes unevenly across these stages: the salary-conversion stage shows the most acute endorsement, while the visibility of relational labor shows the least, consistent with its partial invisibility as a reportable experience.

5.2. Theoretical Integration: The Shared Evaluative Logic

The three facets of the Performance-Recognition Gap are not arbitrary clusters of survey items. Each corresponds to a theoretically specified moment in the evaluation process, and the gendered evaluation literature provides a precise explanatory account for why each moment operates as it does.

The Credibility Tax is grounded in two sets of experimental findings. Steinpreis, Anders, and Ritzke (1999) presented identical credentials under male and female names to faculty evaluators and found that the same document received systematically lower ratings when attributed to a woman, with the pattern reproduced by both male and female evaluators. The mechanism is not individual animus but shared cultural schemas that assign different default assumptions of technical authority to male and female candidates. Uhlmann and Cohen (2005) extended this account by demonstrating that the criteria used to assess merit are not fixed but are constructed post-hoc to favor whichever candidate is perceived as the appropriate choice: when the preferred candidate is male, evaluators shift toward experiential qualifications; when equivalent credentials are attached to a female name, formal credentials are weighted instead. Crucially, evaluators do not recognize that they have moved the criteria. The result is that each individual decision feels principled while the aggregate pattern systematically disadvantages women. Heilman's (2001) lack-of-fit model provides the structural logic: when female role occupancy is perceived as incongruent with the agentic attributes culturally associated with technical leadership, evaluators require more evidence before updating their prior assessment of competence. The Credibility Tax captures the experienced daily effect of operating under this elevated evidentiary standard. Dovidio and Gaertner (2000) add a further precision: their account of aversive bias shows that evaluative discrimination concentrates in the ambiguous middle of assessment,

where neither strong positive nor clearly negative evidence constrains evaluator discretion, and where the gap is most pronounced among evaluators who believe themselves to be fair. In technical environments where almost every judgment about contribution quality is inherently contextual and ambiguous, this creates fertile ground for the Credibility Tax to operate invisibly and continuously.

Transparent Effort is grounded in Castilla and Benard's (2010) meritocracy paradox. In organizations whose cultures explicitly celebrate objective assessment and individual meritocracy, managers award larger performance bonuses to male employees than to equally performing female employees. The mechanism is moral credentialing: the institutional belief that the system is fair reduces the evaluator's impulse to monitor their own judgments for bias. This effect reaches its peak for relational and facilitative contributions, because those contributions occupy no accounting category in evaluation systems built around individually attributed technical outputs. They are real in their effects (teams function, junior colleagues are supported, conflicts are absorbed), but because the formal accounting system has no credit entry for them, the meritocratic logic that fair outcomes follow from fair processes renders their non-registration invisible. The Israeli high-tech sector, whose foundational identity holds that technical output is objective and gender-neutral, creates precisely the conditions under which moral credentialing operates: the belief that "the best code wins" functions as a structural license to not examine what the evaluation system counts as output. Transparent Effort's lower top-box endorsement (24.7%) is consistent with this: invisible labor is also, in part, invisible as a distinct reportable experience.

The Broken Scale is grounded in Castilla's (2008) field evidence from a large service organization, where longitudinal compensation data showed that women and minorities who received identical performance ratings to equivalent male colleagues nonetheless received smaller salary increases at the reward allocation step. The parity established at the assessment stage did not survive the discretionary judgment exercised at conversion. This finding is directly relevant to the present study's salary-comparison items, which returned the construct's highest mean endorsement ($M = 3.26$ for the current-peer comparison) and the second-highest top-box score (35.2%). The qualitative data reinforce this pattern: focus-group participants described evaluation language in which men's contributions were characterised in terms of strategic capability and women's equivalent contributions were framed in communal terms, regardless of the similarity of the underlying output. This vocabulary asymmetry encodes differential evaluation into the performance review itself.

5.3. The Double Bind as the Integrating Logic

The three facets are not independent problems that happen to share a measurement instrument. They are connected expressions of the same evaluative

logic, and that logic is most precisely articulated in Eagly and Karau's (2002) role congruity theory.

Role congruity theory holds that perceived incongruity between the communal attributes ascribed to women and the agentic attributes associated with technical and leadership roles produces two routes to undervaluation. The first operates before any behavior is observed: female role occupancy in a technical or leadership position triggers lower prior assumptions of competence, generating the Credibility Tax before the first contribution has been made. The second operates in response to behavior itself: when women enact the technically assertive, authority-claiming behaviors required for recognition and advancement, they violate the communal norm, triggering what Rudman and Glick (2001) term backlash: negative social and professional consequences for displaying the very behaviors that would be rewarded in a male colleague.

The evaluative logic produces a structural trap. A woman who performs communal, facilitative, and relationship-maintaining work satisfies social expectations but finds that work rendered invisible in evaluation accounting (the Transparent Effort dynamic). A woman who enacts agentic authority, claims credit, and asserts technical leadership meets the standard that recognition requires, but faces heightened scrutiny and surplus evidentiary demands (the Credibility Tax). Both pathways lead to the same downstream register: the Broken Scale, where the accumulated consequences of unequal credibility and unregistered effort converge in the gap between assessed performance and actual compensation.

This integrating framework explains why the Performance-Recognition Gap is not resolved by individual strategy adjustment. No behavioral orientation eliminates the evaluation asymmetry; each produces a different version of the same mechanism. Interview participants articulated this trap directly, describing the double bind as a practical constraint on how they could present themselves and their work. The construct's convergence across the career outcomes examined here, regardless of women's seniority, tenure, or professional background, is consistent with this prediction: the mechanism operates on women's professional experience, not on a specific subset of behavioral strategies.

5.4. Everyday Evaluation and Career Consequences

The four career outcome associations form a temporal arc: a description of how an evaluative mechanism, operating through everyday moments of credit attribution and recognition, accumulates into a career trajectory.

Advancement expectancy ($r = -.530$) shows the strongest correlation in the dataset, exceeding that with observed career advancement ($r = -.494$): it captures the belief that sustained effort will convert into commensurate career progress. This ordering is substantively meaningful. A woman who has encountered

repeated instances in which her contributions were attributed to a male colleague, her relational labor passed unregistered through a performance review, or her salary remained below that of peers with equivalent assessments, may not observe a dramatically stalled career record in the immediate term. The more proximate consequence is a recalibration of expectation: a revised estimate of the probability that future effort will receive different treatment than past effort has received. Advancement expectancy captures this erosion before it manifests as formal exit or visible career stagnation. The association with exit intention ($r = .412$) indicates that by the time women actively consider leaving, the evaluative mechanism may already have been operating long enough to make belief in career convertibility difficult to sustain.

Professional safety perception ($r = -.387$) constitutes a related but distinct downstream consequence. The qualitative data suggest that the evaluation environment described by respondents is one in which expressing disagreement, claiming credit, or asserting technical authority carries reputational risk. When candour is associated with social cost, professional safety becomes a precondition for the very behaviors through which recognition is obtained, creating a secondary constraint on women's visibility that reinforces the primary mechanism.

That these associations survive adjustment for seniority, tenure, and demographic background indicates that the mechanism operates independently of career stage and organizational position. The implication is an early-warning insight: the most visible outcome (formal exit or stalled advancement) may be a trailing indicator of a mechanism that has been operating at the level of everyday evaluation, expectation formation, and professional safety for considerably longer.

5.5. Organizational Implications in Israeli High-Tech

The practical implication of a structural mechanism is that it cannot be resolved by intervening at only one point in the evaluation chain. The Performance-Recognition Gap produces recognizable consequences at three distinct stages, and each represents a site where evaluation accountability can be strengthened.

The highest-leverage intervention point is the earliest promotion transition. McKinsey and Company (2021) document that the steepest advancement attrition is not at the board level but at the first management gate, where informal assessments of technical credibility function as tacit criteria alongside formal performance indicators. In the Israeli high-tech sector, calibration meetings and informal group discussions typically constitute the evaluative moment at which these assessments are formed and recorded. Structuring the calibration process to require that claims about leadership readiness be grounded in specified contribution evidence, and that the attribution of ideas be made

explicit rather than assumed, could reduce the evaluative space in which the Credibility Tax operates most freely. The interview account of assertiveness being applied as a threshold criterion to female candidates alone illustrates precisely the kind of tacit shift that structured evaluation documentation is designed to interrupt.

For Transparent Effort, the practical implication is the provision of an accounting category. Evaluation frameworks that formally include facilitative and relational contributions (team mentoring, onboarding leadership, knowledge transfer, conflict resolution) alongside technical deliverables create a visible credit line for work that currently passes through the system without registration. The absence of such a category is a structural choice that systematically disadvantages the workers who perform that work disproportionately.

The domain effect observed in the adjusted analysis, whereby women in core technical functions report significantly lower career advancement after the Performance-Recognition Gap is controlled, suggests that additional structural barriers may operate in research and development, engineering, and data science environments beyond those the construct captures. These domains carry the highest symbolic and economic capital in the Israeli high-tech ecosystem, and the concentration of evaluation inequity in precisely these roles makes audit of calibration and promotion practices in technical function lines a particularly high-priority organizational concern.

The null finding for military background ($p = .361$) adds important context for the Israeli ecosystem specifically. Elite military technology credentials carry recognized signaling value in hiring and network access; they do not appear to buffer women against recognition asymmetries once they are inside organizational settings. The credential modifies entry conditions; it does not modify the evaluation environment women encounter after entry.

5.6. Limitations

This study offers an initial adapted operationalization of the Performance-Recognition Gap in the Israeli high-tech context rather than a full psychometric validation. The three-facet structure proposed here emerged from reliability analysis and interpretive synthesis, and its stability across independent samples and discriminant validity relative to related constructs are not yet established. The findings should be interpreted in light of the cross-sectional design, which does not support causal inference, and the self-report outcome measures, which may share method variance with the predictor construct.

These boundaries define the scope of the current contribution rather than undermine it.

6. CONCLUSION

6.1. Contributions to the Literature

This study shows that the Performance-Recognition Gap is observable and meaningfully measurable among women in the Israeli high-tech sector. The construct demonstrated strong internal consistency and a coherent three-facet interpretive structure comprising the Credibility Tax, Transparent Effort, and the Broken Scale. Across the study's qualitative and quantitative components, the findings suggest that evaluative credit disparities reflect a broader recognition mechanism operating through everyday organizational practices.

The study also shows that the Performance-Recognition Gap is associated with career outcomes, including lower relative career advancement, lower advancement expectancy, lower professional safety, and higher exit intention. The strongest observed association was with advancement expectancy, suggesting that erosion of progress expectations may emerge before more visible outcomes such as stalled advancement or exit. In adjusted analysis, the construct retained a significant negative association with relative career advancement, indicating that the pattern is not confined to a single aspect of professional experience.

A third contribution is the identification of core technical roles as a context in which evaluative credit disparities appear especially consequential for career advancement. This highlights research and development, engineering, and data science environments as particularly important sites for organizational audit and intervention.

6.2. Future Research

Future research could extend this work through confirmatory factor analysis and convergent validity testing to examine the stability of the three-facet structure across independent samples. Comparative or longitudinal designs would also help clarify how the Performance-Recognition Gap operates across contexts and over time.

References

1) AnitaB.org. (2022) *Technical Equity Experience (TEE) 2022 Report*. Redwood City: AnitaB.org.
2) Castilla, E. J. (2008) 'Gender, race, and meritocracy in organizational careers', *American Journal of Sociology*, 113(6), pp. 1479-1526.
3) Castilla, E. J. and Benard, S. (2010) 'The paradox of meritocracy in organizations', *Administrative Science Quarterly*, 55(4), pp. 543-576.
4) Diehl, A. B., Stephenson, A. L. and Dzubinski, L. M. (2020) 'Measuring the invisible: Development and multiindustry validation of the Gender Bias Scale for Women Leaders', *Human Resource Development Quarterly*, 31(3), pp. 249-280.

5) Dovidio, J. F. and Gaertner, S. L. (2000) 'Aversive racism and selection decisions', *Psychological Science*, 11(4), pp. 315-319.
6) Eagly, A. H. and Karau, S. J. (2002) 'Role congruity theory of prejudice toward female leaders', *Psychological Review*, 109(3), pp. 573-598.
7) Hair, J. F., Black, W. C., Babin, B. J. and Anderson, R. E. (2009) *Multivariate Data Analysis: A Global Perspective*. 7th edn. Upper Saddle River, NJ: Pearson Education.
8) Heilman, M. E. (2001) 'Description and prescription: How gender stereotypes prevent women's ascent', *Journal of Social Issues*, 57(4), pp. 657-674.
9) Israel Innovation Authority. (2025) *State of High-Tech 2025*. Jerusalem: Israel Innovation Authority.
10) McKinsey and Company. (2021) *Women in the Workplace 2021*. New York: McKinsey and Company with LeanIn.Org.
11) McKinsey and Company. (2024) *Women in the Workplace 2024*. New York: McKinsey and Company with LeanIn.Org.
12) PayScale. (2024) *The State of the Gender Pay Gap 2024*. Seattle: PayScale Inc.
13) Rudman, L. A. and Glick, P. (2001) 'Prescriptive gender stereotypes and backlash toward agentic women', *Journal of Social Issues*, 57(4), pp. 743-762.
14) Steinpreis, R. E., Anders, K. A. and Ritzke, D. (1999) 'The impact of gender on the review of the curricula vitae of job applicants and tenure candidates: A national empirical study', *Sex Roles*, 41(7/8), pp. 509-528.
15) Uhlmann, E. L. and Cohen, G. L. (2005) 'Constructed criteria: Redefining merit to justify discrimination', *Psychological Science*, 16(6), pp. 474-480.
16) World Economic Forum. (2022) *Global Gender Gap Report 2022*. Geneva: World Economic Forum.

APPENDIX
APPENDIX A. PERFORMANCE-RECOGNITION GAP: COMPLETE ITEM WORDINGS

The following table presents the exact wordings of the ten items used to derive the scored Performance-Recognition Gap construct, drawn from the adapted Gender Bias Scale for Women Leaders (Diehl et al., 2020). Nine items are retained in the scored construct; one item (Item 8, marked †) was dropped following item-level psychometric diagnostics and is included here for complete instrument documentation. All items were rated on a five-point Likert scale (1 = Strongly Disagree to 5 = Strongly Agree). Reverse-coded items are marked R: for these items, higher agreement reflects a lower level of the barrier experience.

Table A1. Performance-Recognition Gap: Complete Item Wordings by Facet

Facet	Item	Item wording	Notes
Credibility Tax	1	My ideas seem more likely to be taken seriously when a man repeats them	
Credibility Tax	2	My job performance has been scrutinized more closely than that of my male colleagues	
Credibility Tax	3	As a woman, I am expected to be nurturing at work	
Credibility Tax	4	I work harder than my male colleagues for the same credibility	
Transparent Effort	5	At work, I am interrupted by men when I am speaking	
Transparent Effort	6	When I am the only woman in a meeting, I find it difficult to gain support for my ideas	
Transparent Effort	7	It is taken for granted when I help my male colleagues	
Transparent Effort	8	My efforts at creating harmony at work are noticed	R, †
Broken Scale	9	I have made less money than my male counterparts	
Broken Scale	10	I have made less money than men who held my position before me	

Note. Items are numbered sequentially within the construct for presentation; original instrument item numbers are withheld to protect the adapted instrument. R = reverse-coded item (higher agreement indicates lower barrier experience). † = item dropped from the scored construct following item-level diagnostics: corrected item-total r = .19, below the .30 threshold specified by Hair et al. (2009). The item is retained in this table because its content (recognition of relational harmony-maintenance work) is substantively meaningful and is discussed in Section 4.1 as a direction for future investigation. All items adapted from Diehl, Stephenson, and Dzubinski (2020) for the Israeli high-tech context.