

How AI Feedback Shapes Workplace Social Support Perception: A Randomized Controlled Trial in Japan

Yasushi Watanabe

yasushi.watanabe.77@gmail.com

Kyoto University

Masataka Nakayama

Kyoto University

Kosuke Takemura

Shiga University

Yukiko Uchida

Kyoto University

Research Article

Keywords: AI Feedback, Social Support, Positive/Negative Feedback, Gig Workers, Randomized Controlled Trial

Posted Date: August 6th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9606142/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

Although AI (Artificial Intelligence) is increasingly used to provide workplace feedback, little is known about how AI feedback comes to be perceived as social support by recipients. Social support theory has been developed almost exclusively in the context of human-to-human interaction, leaving open the question of how AI's communication characteristics shape the type of support recipients perceive. This study examined whether AI feedback valence—positive (focusing on strengths and successes) versus negative (focusing on areas for improvement and weaknesses)—is perceived as different types of social support.

Methods

A randomized controlled trial was conducted with 109 Japanese gig workers who were randomly assigned to receive either positive or negative feedback from GPT-4o-based chatbots. Participants discussed their own gig work experiences with the chatbot, which then provided feedback based on their accounts. They interacted with their assigned chatbot over two sessions one week apart, with perceived social support (emotional, informational, and instrumental) assessed immediately after each session and at a one-week follow-up.

Results

Positive feedback was perceived more strongly as emotional support than negative feedback, while negative feedback was perceived more strongly as informational support than positive feedback. The timing at which these differences emerged varied by support type: informational support effects were most pronounced immediately after the first session, whereas emotional support effects were largest at the one-week follow-up. No significant condition differences were observed for instrumental support.

Conclusions

The type of social support perceived from AI feedback can be systematically shaped by the design of its communication. Positive AI feedback functions as emotional support, while negative AI feedback functions as informational support, providing a foundation for designing AI feedback systems that align with workers' diverse needs.

Trial registration:

ISRCTN10031278 (<https://doi.org/10.1186/ISRCTN10031278>). Registered 8 May 2026. Retrospectively registered.

Background

Social support in the workplace is widely recognized as a critical factor influencing employee well-being and organizational outcomes [1]. Defined as the tangible and intangible assistance individuals receive from others, social support has been classified into four types based on its content: emotional support (psychosocial assistance such as empathy and care), informational support (provision of information and advice useful for problem-solving), instrumental support (assistance through concrete resources or direct action), and appraisal support (provision of information that aids self-evaluation) [2]. These forms of support have been linked to a range of positive organizational outcomes, including stress reduction, enhanced job performance, and decreased turnover [1].

Traditionally, this body of research has developed almost exclusively in the context of human-to-human interaction. A recent multidisciplinary review identified diverse sources of social support studied across literatures, spanning organizations, supervisors, mentors, coworkers, family members, friends, neighbors, and even online contacts [3]. Yet across this extensive body of research, support providers have been conceptualized as humans or human-constituted institutions, sharing the implicit assumption that support emerges from interpersonal or institutional relationships.

AI as Support Provider

In recent years, however, the integration of AI technology into the workplace has progressed rapidly [4], and situations in which AI rather than humans serves as the social support provider are becoming a reality [5]. Generative AI such as large language models (LLMs) in particular is capable of human-like linguistic communication, suggesting the potential to function as colleagues, supervisors, or external coaches—roles traditionally associated with the provision of social support. Recent studies provide evidence for AI's potential to function in each of these roles. As coworkers, AI systems have been shown to act as "cybernetic teammates," with individuals working with AI achieving performance levels comparable to human-human teams in collaborative knowledge work [6]. As supervisors providing performance feedback, AI has demonstrated unique advantages: workers were more receptive to negative feedback when delivered by AI rather than human managers, particularly among those concerned about losing face [7]. As external coaches, AI chatbots have been shown to facilitate goal attainment at rates comparable to those achieved through human coaching [8].

While research on AI as a source of social support is emerging, the question of how AI's communication characteristics shape the specific type of support perceived has received little systematic attention. Given the human-like conversational capabilities of contemporary generative AI, it is readily conceivable that AI can be perceived as providing informational support when it offers useful information, appraisal support when it offers evaluative feedback, or instrumental support when it directly assists with problem-solving (e.g., writing programming code). These functions—providing information, evaluation, or concrete assistance—align naturally with what AI systems can technically deliver. Emotional support, however, presents a different question: it is not yet clear whether—and through what kind of AI communication—recipients perceive AI as providing emotional support.

Whether AI can also be perceived as providing emotional support remains a matter of debate. On one hand, recent studies suggest that people can perceive emotional support from AI: AI-generated messages have been found to convey emotional content that recipients evaluate as helpful and authentic [9] and a recent meta-analysis of generative AI chatbots in mental health contexts demonstrated that socially-oriented chatbots—those designed to provide emotional validation and empathy—are more effective in reducing mental health issues than task-oriented ones [5]. On the other hand, concerns remain about whether emotional support can genuinely emerge in the absence of an interpersonal relationship. When recipients are aware that a message comes from AI rather than a human, their evaluations of supportive messages decline [10, 11], and AI feedback has been shown to be experienced with greater social distance than human feedback [12]. The simulated empathy of AI may also undermine perceived sincerity and authenticity [13]. Given these conflicting findings, careful empirical investigation is needed to clarify whether—and through what specific forms of communication—AI can be perceived as providing emotional support. Identifying which AI communication characteristics lead recipients to perceive emotional versus other types is essential for understanding how AI fits into the social support landscape and for the deliberate design of AI communication.

Feedback and Social Support

To investigate this question, the present study focuses on feedback as a key form of AI communication. In human-to-human work environments, workplace feedback inherently overlaps with several functions of social support [2]. It often provides advice and guidance for problem-solving (informational support), conveys evaluative information about performance (appraisal support), and may include suggestions for concrete actions (instrumental support, the provision of tangible aid). Depending on its content, feedback can also convey care and affirmation (emotional support, the provision of psychosocial assistance such as empathy and caring). This functional overlap suggests that AI feedback may be perceived as different types of social support depending on its communication characteristics.

Among the various communication characteristics of feedback, *valence*—whether feedback emphasizes positive aspects (strengths and successes) or negative aspects (areas for improvement)—has received particularly extensive attention in human feedback research. Decades of research have accumulated evidence not only that feedback valence shapes its effects on recipients [14], but also that these effects vary systematically with boundary conditions—for example, cultural orientation [15], learning contexts [16], and individual personality [17]. This rich theoretical foundation makes feedback valence a productive framework for examining AI feedback: established findings from human feedback research can be tested against AI feedback, allowing researchers to identify whether the same patterns hold in AI contexts or whether meaningful differences emerge. Recent

studies have begun applying the feedback valence framework to AI feedback [18, 19]. The present study extends this line of research by examining how feedback valence shapes the type of social support recipients perceive from AI communication.

Hypothesis Development and Current Study

Building on this framework, we now consider how AI feedback valence may shape the type of social support recipients perceive. AI-based feedback systems differ from human feedback in a fundamental way: their communication style is determined by prompt instructions and can be deliberately configured to deliver different content. Feedback valence is one such fundamental dimension of design that may shape how recipients perceive the support they receive.

Positive feedback is designed to recognize achievements, highlight strengths, and affirm the recipient's competence. These features—conveying recognition, affirmation, and positive evaluation of the recipient—closely correspond to *emotional support*, defined as the provision of psychosocial assistance such as empathy and caring [2]. Although the perceived social distance from AI may raise concerns about its capacity to convey empathy [12], the present study's work-reflection context centers on the recognition and affirmation of participants' own concrete achievements. Under such conditions, the affirmation itself can function as emotional support, and AI feedback emphasizing positive aspects of the recipient's work may therefore be perceived as carrying emotional support.

Negative feedback, in contrast, is designed to identify areas for improvement and to provide guidance for problem-solving. These features—specifying weaknesses and offering actionable suggestions—closely correspond to *informational support*, defined as the provision of advice and information useful for problem-solving [2]. Moreover, the social distance from AI may facilitate the reception of critical content: such distance can reduce the self-concept threat typically associated with negative feedback [7], allowing the feedback to be received more straightforwardly as informational support. AI feedback emphasizing areas for improvement may therefore be perceived as carrying informational support.

Based on the theoretical reasoning outlined above, we propose the following hypotheses:

Hypothesis 1

Positive feedback from AI will be perceived as carrying emotional support more strongly than negative feedback.

Hypothesis 2

Negative feedback from AI will be perceived as carrying informational support more strongly than positive feedback.

These hypotheses focus on emotional and informational support among the four types of social support proposed by House [2]. For the remaining two types—instrumental and appraisal support—we do not formulate directional hypotheses, but the reasons differ between the two. *Instrumental support*, defined as the provision of tangible aid [2], may be perceived to some degree from feedback that includes highly concrete, actionable suggestions, which can give recipients a sense of being supported through the problem-solving process. By definition, however, instrumental support is somewhat distant from communicative acts such as feedback. We therefore include instrumental support in our analyses but do not predict valence-related differences. *Appraisal support*, defined as the provision of evaluative information that aids self-evaluation [2], is not analyzed in the present study. The AI chatbot used in our experiment was developed independently from participants' actual workplaces and therefore lacked access to specific performance standards or evaluation criteria. Under this experimental design, the AI feedback could not function as appraisal support, and we accordingly omit it from analysis.

Methods

Participants

Participants were recruited in May 2025 through Lancers, a Japanese crowdsourcing platform. We selected gig workers as the study population because gig work is a context where AI-based feedback is increasingly relevant: work assignment and

performance evaluation are already performed by algorithms in this domain [20], and AI-based feedback is likely to be implemented earlier here than in traditional occupations.

A total of 241 individuals responded to the screening survey. Respondents were excluded if they: provided duplicate responses ($n = 28$), engaged in gig work less than once per week ($n = 47$), completed the survey in substantially shorter time than expected ($n = 8$), had high rates of missing data ($n = 26$), or failed attention checks ($n = 44$). Because some respondents met multiple criteria, the total number of unique exclusions was 109, resulting in 132 participants eligible for randomization.

The 132 participants selected for the experiment were stratified by three variables—age group, sex, and type of gig work (warehouse/logistics, delivery, or other)—and then randomly assigned to one of two conditions using a computer-generated random sequence in R. This stratified randomization procedure resulted in 68 participants assigned to the positive feedback condition and 64 to the negative feedback condition.

Of the 132 randomized participants, 23 (12 in the positive condition, 11 in the negative condition) did not participate in any feedback session and were excluded from the analyses. The remaining 109 participants (56 in the positive condition, 53 in the negative condition) provided data for at least one of the three measurement time points and were included in the primary analyses. Among these 109 participants, 101 (51 in the positive condition, 50 in the negative condition) provided complete data across all three time points (retention rate: 76.5%). A CONSORT flow diagram detailing participant flow is presented in Fig. 1.

This study was retrospectively registered in the ISRCTN registry (ISRCTN10031278) on 8 May 2026. The study materials and analysis code are publicly available on the OSF (see the Availability of data and materials).

Flowchart of Randomized Controlled Trial

Table 1 presents the demographic and work-related characteristics of participants in the complete-data subsample ($n = 101$). The mean age of the sample was approximately 40 years (positive condition: $M = 42.35$, $SD = 9.79$; negative condition: $M = 39.51$, $SD = 8.38$). The majority of participants were male (positive condition: 76.5%; negative condition: 74.0%). On average, participants engaged in gig work for approximately 10 hours per week (positive condition: $M = 10.73$, $SD = 11.26$; negative condition: $M = 9.92$, $SD = 9.22$). Education levels were relatively high, with the majority having completed at least a university education. The most common types of gig work were basic manual work (positive: 47.1%; negative: 56.0%) and food delivery (positive: 29.4%; negative: 30.0%).

To verify the effectiveness of randomization, between-group comparisons were conducted for all baseline characteristics, including variables not used for stratification. Independent samples t -tests (Welch's t -test) were used for continuous variables, and chi-square tests or Fisher's exact test (when expected cell counts were less than 5) were used for categorical variables. No statistically significant differences were observed between the positive and negative feedback conditions on any variable (all p -values $> .05$). These results indicate that randomization was successful.

The study protocol received ethics approval (see Declarations). Participants provided informed consent before participation, including consent for data use in academic publications, and were explicitly informed that the study involved interactions with AI chatbots. Participants received 100 JPY (approximately 0.65 USD) for completing the screening survey and an additional 900 JPY (approximately 6 USD) for completing both the feedback sessions and the post-survey.

Table 1
Demographic and Work-related Characteristics: Comparison Between Negative and Positive Groups

Variable	Negative group (n = 50)	Positive group (n = 51)	Statistical Comparison
Average Age	39.51 (8.38)	42.35 (9.79)	$t(95.04) = -1.54, p = .13$
Gender			$\chi^2(1) = 0.06, p = .82$
Male	74.0% (37)	76.5% (39)	
Female	26.0% (13)	21.6% (11)	
Education level			$\chi^2(4) = 2.60, p = .63$
Junior high school graduate	2.0% (1)	0.0% (0)	
High school graduate	22.4% (11)	15.7% (8)	
Vocational school graduate	12.2% (6)	19.6% (10)	
University graduate	57.1% (28)	56.9% (29)	
Graduate school	6.1% (3)	7.8% (4)	
Weekly working hours	9.92 (9.22)	10.73 (11.26)	$t(98.09) = -0.39, p = .70$
Weekly income (JPY)	17092 (27,667)	15373 (18,161)	$t(82.46) = 0.37, p = .72$
Gig work experience (months)	2.78 (1.66)	2.72 (1.50)	$t(97.00) = 0.17, p = .86$
Type of gig work (multiple response)			
Food delivery	30.0% (15)	29.4% (15)	$\chi^2(1) = 0.00, p = 1.00$
Basic manual work	56.0% (28)	47.1% (24)	$\chi^2(1) = 0.49, p = .48$
Restaurant work	14.0% (7)	17.6% (9)	$\chi^2(1) = 0.05, p = .82$
Event work	16.0% (8)	9.8% (5)	$\chi^2(1) = 0.40, p = .53$
Care assistance	4.0% (2)	7.8% (4)	Fisher's exact, $p = .68$
Retail work	10.0% (5)	7.8% (4)	Fisher's exact, $p = .74$
Package delivery/Moving	12.0% (6)	9.8% (5)	$\chi^2(1) = 0.00, p = .97$

Note: For continuous variables, n indicates the number of valid responses, and values are presented as Mean (SD). For categorical variables, values are presented as % (n); percentages are based on valid responses, as some data were missing. Continuous variables (Age, Weekly working hours, Weekly income, Gig work experience) were compared using independent samples *t*-tests (Welch's *t*-test). Categorical variables (Gender, Education level, Type of gig work) were compared using chi-square tests or Fisher's exact test when expected cell counts were less than 5.

Sample Size Analysis

The sample size was determined through simulation-based power analysis using linear mixed-effects models. To detect the main effect of group (positive versus negative feedback), which is the primary objective of this study, power was evaluated using a mixed-effects model that controlled for individual differences as random intercepts. Based on prior research on AI feedback [18, 19] reporting small to medium effects, we set Cohen's $d = .35$ for the effect size.

For the intraclass correlation coefficient (ICC), we referred to a previous study on the effects of AI feedback on occupational self-efficacy [18], which reported an ICC of approximately .90. However, we expected greater variability in the present study because, unlike self-efficacy—a relatively stable self-perception—perceived social support involves evaluation of an external agent, which is likely to fluctuate more across measurements. Accordingly, we adopted a more conservative estimate of ICC = .60. With $\alpha = .05$ and target power of 80%, 500 simulations were conducted, demonstrating that 60 participants per group (120 total) would yield 82% power. Accounting for attrition, the study was designed with a target of approximately 65 participants per group.

Of the 132 participants randomized, 109 (56 in the positive group, 53 in the negative group) provided data for at least one measurement time point and were included in the primary analyses; 101 of these (51 in the positive group, 50 in the negative group) completed all three measurement time points. The adequacy of this realized sample size given the observed variance structure is examined in the Results section.

Experiment Design

This study employed a longitudinal randomized controlled trial design with two groups (positive feedback versus negative feedback) and three measurement time points: immediately after Session 1 (T1), immediately after Session 2 (T2), and post-intervention (POST).

The study's measurement schedule was designed to align with realistic patterns of work reflection: a weekly review of one's own work at the end of each week. Accordingly, the two feedback sessions were spaced one week apart, providing participants with two opportunities to engage with AI feedback at a naturalistic interval. Two sessions, rather than one, were used to capture both initial reactions to AI feedback and subsequent reactions when participants had become familiar with the format. We considered that participants' first encounter with AI feedback might elicit reactions shaped partly by novelty, whereas repeated exposure could yield reactions less influenced by such effects. The AI was configured to begin each session without memory of prior sessions, ensuring that any change in perception across sessions reflected participants' own accumulating experience rather than the AI's adaptation to the participant. The post-intervention measurement was taken one week after the second session, matching the interval between sessions. This allowed us to assess whether the perceived support patterns persisted across one feedback-free interval of comparable length.

AI Feedback Intervention

Two AI chatbots with different feedback styles were developed using OpenAI's GPT-4o. The specific prompts for each chatbot were created by making minor modifications to prompts from prior research on AI chatbot feedback [18, 19], adapting questions about work situations to the gig work context. Each AI chatbot was implemented on an independent website, and participants interacted with the AI chatbot assigned to their condition via a web browser. The two conditions differed in how the AI delivered feedback during the chat session.

In the positive feedback condition, the AI was instructed to focus on what the participant did well and on their strengths, knowledge, abilities, and talents. The system prompt explicitly directed the AI to praise these elements with strong expressions, providing examples such as "Persevering through a difficult situation reflects your strong perseverance" and "Being able to bring the team together speaks to the strength of your leadership." When participants spoke about failures or shortcomings, the AI was instructed to find positive elements within them and continue the conversation accordingly. The prompt explicitly directed the AI not to point out the participant's weaknesses or shortcomings.

In the negative feedback condition, the AI was instructed to focus on the knowledge, abilities, and experiences the participant should develop. The system prompt explicitly directed the AI to identify weaknesses and encourage improvement, providing examples such as "To navigate difficult situations, you may need to strengthen your thinking and concentration abilities" and "To take on a role of bringing the team together, you may need to enhance your leadership skills." When participants spoke about successes or strengths, the AI was instructed to find improvement points within them and continue the conversation accordingly. The prompt explicitly directed the AI not to praise the participant or offer positive remarks.

Other elements of the system prompts were identical across conditions, including the AI's role as a mentor supporting weekly work reflection, conversation duration (approximately 15 minutes), and stylistic instructions (e.g., keeping responses concise, limiting each response to one main question, alternating between questions and suggestions). The full prompt texts in both the original Japanese and English translation are available in the OSF repository (see Availability of data and materials).

Participants were blinded to conditions and were not informed of the existence of other experimental conditions. The experimenter was not involved in the feedback session process; all interactions proceeded autonomously between the AI and participants.

Procedure

The experiment was conducted entirely online and proceeded as follows (see Fig. 2 for an overview).

Screening survey

Participants completed a web-based screening survey lasting approximately 15 minutes. This survey served to screen participants and obtain baseline information, including stratification variables for group assignment. Participants reported demographic information, gig work engagement (platforms used, types of tasks performed, years of experience, etc.), and baseline measures (average weekly days of gig work, working hours, weekly income, etc.).

Feedback sessions

After completing the screening survey, eligible participants received an invitation to the AI feedback session for their assigned condition. To balance standardization of session timing with high participation rates, participants were given a four-day window (Friday through Monday) each week to complete their session. Session 1 (T1) was conducted approximately two weeks after completion of the screening survey, and Session 2 (T2) was conducted one week after T1.

Each feedback session lasted 15 to 30 minutes and consisted of three steps. First, participants reported their gig work activity for the week (days worked and earnings). Second, in the AI chat session, the AI chatbot asked about specific tasks performed during the past week, posed follow-up questions about behaviors and thoughts during work, and delivered feedback based on the participant's responses. Third, after the chat session, the AI directed participants to a web-based questionnaire where they completed a manipulation check and reported perceived social support from the AI (emotional, informational, and instrumental).

Post-survey (Post)

Approximately one week after Session 2 (T2), participants completed the same measures to assess whether the perceived support patterns persisted during the period without further feedback.

Experimental Procedure and Measurement Schedule

Measures

Perceived social support from AI: The Brief Workplace Social Support Scale [21] was used to measure three dimensions of social support: informational support, instrumental support, and emotional support. Each dimension consists of two items. Participants were asked to what extent they felt they had received support during the session with the AI chatbot, rating each item on a 5-point Likert scale (1 = almost never, 5 = very much). Sample items include "understands and acknowledges you" for emotional support, "gives advice for solving problems" for informational support, and "works through things with you" for instrumental support. Internal consistency across the three measurement time points ranged from $\alpha = .81-.83$ (inter-item $r = .68-.71$) for informational support, $\alpha = .62-.74$ ($r = .45-.59$) for instrumental support, and $\alpha = .71-.89$ ($r = .55-.81$) for emotional support. Discriminant validity among the three types of social support was supported by the Fornell-Larcker criterion: AVE (Average Variance Extracted) values ranged from .72-.90, exceeding the squared inter-construct correlations (.02-.60) at all time points.

Manipulation check

Participants were asked to evaluate the content discussed during the session with the AI. Following the method used by prior AI chatbot feedback studies [18, 19], participants rated the session on a 7-point scale ranging from 1 (mainly discussed lacking knowledge and problems to be solved; corresponding to the negative feedback condition) to 7 (mainly discussed strengths and things done well; corresponding to the positive feedback condition).

Results

Manipulation Check

Independent samples *t*-tests at each time point revealed that the positive feedback condition showed significantly higher values (i.e., greater perception that the dialogue focused on strengths and successes) than the negative feedback condition at all

measurement time points. At T1, the positive condition ($M = 5.48$, $SD = 1.34$) was significantly higher than the negative condition ($M = 3.38$, $SD = 1.93$), $t(104) = -6.47$, $p < .001$, with a large effect size (Cohen's $d = 1.26$). At T2, although the effect size was somewhat smaller (Cohen's $d = 0.68$), the between-group difference remained clear (negative condition: $M = 3.94$, $SD = 2.01$; positive condition: $M = 5.23$, $SD = 1.78$), $t(102) = -3.44$, $p < .01$. At Post, the positive condition ($M = 5.43$, $SD = 1.37$) again showed significantly higher values than the negative condition ($M = 3.42$, $SD = 1.70$), $t(104) = -6.68$, $p < .001$, with a large effect size (Cohen's $d = 1.30$).

These results indicate that participants in the positive feedback condition perceived the discussion as focusing primarily on strengths and successes, whereas participants in the negative feedback condition perceived the discussion as focusing primarily on challenges and areas for improvement. Given that medium to large effect sizes were observed at all time points, the experimental manipulation was deemed to have functioned appropriately.

Descriptive Statistics

Means, standard deviations, and intercorrelations among the perceived social support variables are presented in Table 2. First, mean scores at all time points exceeded the scale midpoint of 3.0, and emotional support consistently averaged above 4.0, indicating that participants perceived substantial social support from AI feedback overall. Second, intercorrelations among the three support types within the same time point were moderate rather than high (ranging from .14 to .63), suggesting that the three types were perceived as relatively distinct dimensions. Correlations of the same support type across time points were also moderate (e.g., $r = .55$ – $.66$ for informational support), consistent with the interpretation that participants' perceptions reflected each specific session rather than a stable trait-like evaluation.

Sensitivity Power Analysis

We conducted a sensitivity power analysis for two reasons. First, the realized sample size for the primary analyses ranged from $n = 53$ – 56 per group, slightly smaller than the recruitment target due to attrition; we computed the sensitivity power using the complete-data subset ($n = 50$ per group) as a stricter criterion. Second, the ICC = .60 used in the a priori power analysis was extrapolated from a study on a different construct, and the actual variance structure of the present outcomes was not known at the design stage. We therefore estimated null models (random intercepts only) to obtain the observed ICCs for each support type and used these values to recompute the simulated power.

The observed ICCs were .610 for informational support, .583 for instrumental support, and .461 for emotional support, close to or somewhat lower than the assumed value of .60. Using these ICCs and the anticipated effect size ($d = .35$), simulated power for $n = 50$ per group was 80.4%, 81.0%, and 81.0% for informational, instrumental, and emotional support, respectively, indicating adequate power to detect the anticipated effect for all three outcomes.

Table 2
Means, Standard Deviations, and Correlations Among Perceived Social Support Variables

Variable	M	SD	1	2	3	4	5	6	7	8	9
Informational Support											
1. Time1	3.89	0.90	—								
2. Time2	3.90	0.94	0.63***	—							
3. Post	3.95	0.92	0.55***	0.66***	—						
Instrumental Support											
4. Time1	3.94	0.78	0.63***	0.54***	0.52***	—					
5. Time2	3.90	0.85	0.48***	0.78***	0.65***	0.62***	—				
6. Post	3.87	0.93	0.51***	0.52***	0.57***	0.51***	0.62***	—			
Emotional Support											
7. Time1	4.33	0.87	0.20*	0.14	0.14	0.34***	0.30**	0.27**	—		
8. Time2	4.31	1.00	0.26**	0.43***	0.19	0.34***	0.53***	0.56***	0.43***	—	
9. Post	4.27	0.98	0.22*	0.36***	0.15	0.30**	0.36***	0.48***	0.30**	0.63***	—
Note. * $p < .05$. ** $p < .01$. *** $p < .001$.											

Main Analysis: Effects of AI Feedback on Social Support Perception

To examine whether positive and negative AI feedback styles differently affect perceptions of social support, we fitted separate linear mixed-effects models for each support type (emotional, informational, and instrumental) using a two-step hierarchical approach.

In the first step (Model 1), we fitted a main-effects model including condition (effect-coded: positive = 1, negative = - 1) and time (T1, T2, and Post as categorical variables) as fixed effects, with random intercepts for participants. This model tests whether perceived support differs between conditions across all time points (Table 3). The condition effect was significant for informational support ($B = - 0.153$, $SE = 0.075$, $t = - 2.04$, $p = .044$, $d = 0.34$), indicating that the negative feedback condition perceived higher informational support than the positive feedback condition. For emotional support, the condition effect was marginally significant ($B = 0.141$, $SE = 0.073$, $t = 1.95$, $p = .054$, $d = 0.30$), with the positive feedback condition showing higher perceived emotional support. No significant condition effect was found for instrumental support ($B = - 0.021$, $SE = 0.070$, $t = - 0.29$, $p = .770$, $d = 0.05$).

Table 3
Linear Mixed-Effects Model Results for Perceived Social Support (Model 1: Main Effects Model)

	Informational				Instrumental				Emotional			
	B	SE	t	p	B	SE	t	p	B	SE	t	p
Fixed effects												
Intercept	3.88***	0.09	44.3	< .001	3.92***	0.08	47.4	< .001	4.33***	0.09	47.4	< .001
Time (T2)	0.03	0.08	0.43	.67	−0.02	0.08	−0.23	.82	−0.03	0.10	−0.33	.75
Time (Post)	0.09	0.08	1.07	.28	−0.04	0.08	−0.50	.62	−0.06	0.10	−0.59	.55
Condition (P = 1, N = − 1)	−0.15*	0.08	−2.04	.044	−0.02	0.07	−0.29	.770	0.14†	0.07	1.95	.054
Random effects												
τ ₀₀ (Intercept)	0.50				0.43				0.40			
σ ² (Residual)	0.33				0.31				0.49			
ICC	.60				.58				.45			
Note. Condition was effect-coded (Positive = 1, Negative = − 1). Time was dummy-coded with T1 as the reference category. ICC = intraclass correlation coefficient.												

[†] $p < .10$. * $p < .05$. *** $p < .001$.

As outlined in the Methods, the study's three-time-point design was intended to capture both initial and longer-term patterns of perceived support, regardless of whether the overall condition $\times$ time interaction reached significance. We therefore additionally fitted a model including the condition $\times$ time interaction (Model 2) and computed estimated marginal means (EMMs) at each time point (Fig. 3) to descriptively characterize the temporal patterns. Although the interaction did not significantly improve model fit (informational: $\chi^2(2) = 5.76$, $p = .056$; instrumental: $\chi^2(2) = 0.74$, $p = .69$; emotional: $\chi^2(2) = 1.34$, $p = .51$), the EMMs allow comparison of the predicted pattern at each measurement occasion. Full model details are reported in Table S1 in the Supplementary Materials on OSF.

Across all time points, mean scores exceeded the scale midpoint of 3.0 for all three support types in both conditions. In the positive condition, mean scores ranged from 3.63 to 3.85 for informational support, 3.87 to 3.91 for instrumental support, and 4.40 to 4.47 for emotional support. In the negative condition, mean scores ranged from 3.98 to 4.14 for informational support, 3.90 to 3.98 for instrumental support, and 4.07 to 4.21 for emotional support.

Descriptively, for informational support, the negative feedback condition showed a numerically larger between-condition difference at T1 (negative: $M = 4.14$; positive: $M = 3.64$; $d = 0.55$, $p = .005$), but this difference was not significant at T2 ($d = 0.14$, $p = .474$) or Post ($d = 0.31$, $p = .111$). This pattern suggests that the informational support effect identified in Model 1 was primarily driven by the first session. For instrumental support, no significant differences were found at any time point (all $ps > .49$). For emotional support, no significant between-condition differences were observed at T1 ($d = 0.25$, $p = .200$) or T2 ($d = 0.21$, $p = .276$), although the positive feedback condition consistently showed higher scores. At Post, one week after the final session, the positive feedback condition showed a larger between-condition difference than at earlier time points (positive: $M = 4.47$; negative: $M = 4.07$; $d = 0.43$, $p = .027$). This delayed emergence of the emotional support effect may partly account for why the overall condition effect in Model 1 was only marginally significant—the effect was diluted by the earlier non-significant time points.

Estimated Marginal Means of Perceived Social Support by Condition and Time Point (Model 2)

Note

Error bars represent 95% confidence intervals. Solid lines with circles indicate the positive feedback condition; dashed lines with triangles indicate the negative feedback condition. Significance markers denote between-condition differences at each time point. $*p < .05$. $**p < .01$.

Discussion

This study examined how positive versus negative AI feedback is perceived as different types of social support by recipients. The results supported both hypotheses: positive feedback was perceived more strongly as emotional support than negative feedback, and negative feedback was perceived more strongly as informational support than positive feedback. Notably, the temporal patterns of these effects appeared to differ between the two support types: descriptively, differences in informational support were most pronounced immediately after the first session, whereas differences in emotional support were larger at the one-week follow-up than at earlier sessions. Below, we discuss substantive interpretations of these findings, theoretical implications, and practical implications.

The findings for emotional support speak directly to the central question of this study: how does AI-delivered feedback come to be perceived by recipients? Although AI is generally perceived as more socially distant than human communicators [12], the present results show that when AI recognizes what the recipient has actually done well in their own work, recipients perceive this recognition as emotional support more strongly than when AI emphasizes areas for improvement. Notably, mean scores for emotional support in the positive feedback condition exceeded 4 on the 5-point scale at all time points, indicating that participants perceived substantial emotional support from the AI when it delivered positive feedback.

Why was the positive feedback from AI perceived as emotional support despite this social distance? A theoretically meaningful interpretation of this finding draws on Bandura's [22] work on self-efficacy. Bandura emphasized that mastery experiences—one's own actual successes—exert a particularly strong influence on self-efficacy, more so than verbal persuasion from others. Building on this insight, when affirmation and recognition are anchored in one's own concrete successes, they may carry an inherent sense of credibility. In the present study, the AI's affirming feedback was grounded in participants' own recent work activities that they had themselves identified. Receiving recognition for these concrete experiences may have evoked a sense of being validated, even though the source was a chatbot that participants knew to be a program. This interpretation suggests that AI feedback can be perceived as emotionally supportive when the interaction is designed around the recipient's own experiences.

This perception of being recognized appears to leave a durable impression. In the positive feedback condition, perceived emotional support did not decline at the one-week follow-up; rather, the between-condition difference was largest at this point (descriptively, $d = 0.43$). The recognition of one's own successful experiences—affirmed through the AI interaction—may have settled into participants' memory of the experience, sustaining the perception of having received emotional support over the subsequent week without further feedback.

The pattern observed for informational support similarly speaks to how AI-delivered feedback comes to be perceived. When AI delivered feedback emphasizing what the recipient lacked or needed to develop based on their own work, recipients perceived this as informational support more strongly than when AI delivered positive feedback. Mean scores for informational support in the negative feedback condition were around 4 on the 5-point scale, indicating that participants perceived the AI's negative feedback as substantive information for problem-solving.

Why was AI-delivered negative feedback perceived as informational support, despite the well-known difficulty of delivering negative feedback effectively? Negative feedback has typically been characterized as difficult to deliver effectively in interpersonal communication research, because recipients often experience it as a threat to the self, which reduces the feedback's effectiveness as task-relevant information [23]. One possible interpretation of the present finding is that AI as a feedback source mitigates this self-threat dynamic. The social distance between AI and human recipients [12]—the same characteristic that might otherwise limit the perception of emotional support—may here serve a facilitative role: when the source is AI rather than a human supervisor or colleague, recipients may be released from the need for self-protective interpretation that human-delivered negative feedback often invites. This released cognitive space may have allowed participants to engage with the negative feedback content as substantive

information for improvement, consistent with prior findings that AI-delivered negative feedback can mitigate apprehensions associated with criticism [7].

The between-condition difference in informational support was less pronounced in the second session and at the one-week follow-up. This pattern is open to multiple interpretations, but one possibility relates to a feature of our experimental design: the AI was configured without memory of prior sessions. Although participants likely raised different work experiences across sessions, the absence of cross-session memory may have limited the AI's ability to build on prior conversations and provide increasingly tailored guidance.

Theoretical Contributions

The present study makes two theoretical contributions: one to social support research and one to research on AI feedback.

Contribution to social support research. Social support theory has been developed almost exclusively in the context of human-to-human interaction, with prior research focused on human or human-constituted sources—supervisors, coworkers, family, friends, external coaches, and organizations [3]. These sources all share some form of social role or relational history with the recipient. The present study extends the scope of social support research by positioning AI as a new type of support provider and examining how it operates in this role. Critically, the study addresses how the type of social support perceived from AI varies systematically with its communication style. The demonstration that the type of perceived support can be deliberately shaped through communication design—even in the absence of any pre-existing relationship between recipient and provider—opens a new theoretical direction for social support research.

Contribution to AI feedback research. A growing body of research has examined the effects of AI feedback on workplace outcomes such as performance, self-efficacy, and engagement [7, 8, 18, 19]. These studies have established that AI feedback can produce meaningful effects, but the perceptual processes through which AI feedback comes to influence recipients have remained understudied. The present study contributes to this domain by introducing perceived social support as an intermediate variable through which AI feedback may operate. By demonstrating that AI feedback valence systematically shapes the type of social support perceived—positive feedback as emotional support, negative feedback as informational support—the study provides a foundation for examining how these perceptual differences may mediate the effects of AI feedback on downstream outcomes.

Practical Implications

The present study has practical implications for the design of AI feedback systems in workplace contexts, particularly for gig work where algorithmic management is already established. The findings suggest that AI feedback can function as social support, with the type of support perceived shaped by the design of the feedback. By configuring AI to deliver feedback emphasizing achievements and strengths, designers can create AI systems perceived as providing emotional support; by configuring AI to deliver feedback identifying areas for improvement, designers can create AI systems perceived as providing informational support. This deliberate alignment between communication design and intended support type offers a foundation for designing AI feedback systems that meet specific workplace needs.

Limitations and Future Research

This study has several limitations that point to directions for future research.

First, the sample was limited to gig workers in Japan. We targeted gig workers precisely because algorithmic platforms already manage their work, making AI feedback a natural next step in this context. However, workers who may benefit from AI feedback are not limited to gig workers. As algorithmic management becomes more widespread and the boundaries between organizations and external workers blur, AI feedback implementation may extend to other occupations, including conventional office work. Whether the present findings generalize to other work contexts remains a question for future research. Similarly, given that prior research has demonstrated cultural differences in how positive and negative feedback are perceived [15], the generalizability of these findings to non-Japanese cultural contexts also warrants further investigation.

Second, this study compared two conditions in which the AI delivered consistently positive or consistently negative feedback. In actual feedback situations, positive and negative elements often coexist within a single interaction. The clean separation of valences in the present design allowed us to identify how each communication style is associated with each support type, but how social support perception is formed under mixed conditions remains an open question. Future research should examine mixed feedback conditions to understand how recipients perceive social support when both valences are combined.

Third, the AI chatbot in the present study was configured to operate without memory of prior sessions. While this design choice allowed us to ensure that observed differences across sessions reflected participants' accumulating experience rather than the AI's adaptation, it may have limited the AI's capacity to provide increasingly tailored guidance over time. The diminishing between-condition differences in informational support across sessions may partly reflect this design feature. Future research should examine how AI memory retention shapes the perception and persistence of social support.

Fourth, the present study focused on perceived social support and did not examine downstream outcomes such as work performance, well-being, or behavioral change. Although our findings suggest that AI feedback is perceived as different types of social support depending on its valence, whether and how these perceptual differences translate into outcomes remains a question for future research. Research that connects perception-level findings to outcome-level effects would help clarify the practical impact of AI feedback design.

Fifth, the experimental period was limited to three weeks, with a one-week follow-up. Whether the perceived support patterns observed here persist, strengthen, or diminish with prolonged AI use remains unclear and warrants longer-term investigation.

Finally, the AI chatbot in this study was implemented as an independent tool, separate from any organizational platform. In actual workplace contexts, AI feedback would likely be perceived as representing the organization or platform that deploys it, and this perception of institutional representation could substantially alter how workers receive AI feedback. Furthermore, the AI's independence from participants' actual workplaces meant that it lacked access to organizational performance standards or evaluation criteria, which is why appraisal support fell outside the scope of the present study. Integrating AI feedback systems into organizational platforms would not only change the relational positioning of the AI but also enable the examination of appraisal support—how AI-delivered evaluative feedback against organizational standards is perceived as social support. Future research should examine how social support perception, including the perception of appraisal support, changes when AI is integrated into organizational systems rather than operating as an independent tool.

Conclusion

This study demonstrates that AI can function as a source of differentiated social support in the workplace, with the type of support perceived shaped by the design of AI communication. Feedback valence systematically influenced how recipients perceived the support: positive feedback was perceived as emotional support, while negative feedback was perceived as informational support. As AI-based feedback systems become increasingly prevalent in workplaces, understanding how their communication characteristics shape recipients' perceptions of social support will be essential for designing systems that align with workers' diverse needs.

Abbreviations

AI
Artificial Intelligence
AVE
Average Variance Extracted
CONSORT
Consolidated Standards of Reporting Trials
EMMs
Estimated Marginal Means
GPT
Generative Pre-trained Transformer

ICC
Intraclass Correlation Coefficient
LLMs
Large Language Models
OSF
Open Science Framework

Declarations

Ethics approval and consent to participate

This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the Ethics Committee of the Psychological Science Unit, Kyoto University (Approval No. 6-P-1). Informed consent was obtained from all individual participants included in the study. Participants provided consent by reading the information sheet and clicking to proceed with the online survey.

Consent for publication

Not applicable.

Competing interests

The authors have no relevant financial or non-financial interests to disclose.

Funding

The authors received support from [MASKED FOR REVIEW] for the submitted work.

Author Contribution

All authors contributed to the study conception and design. YW conceptualized the research design, performed the data analysis, and wrote the original draft of the manuscript. MN provided guidance on the research design and analytical approaches. KT and YU contributed to the interpretation of results and critically revised the manuscript. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

Data Availability

The datasets, analysis code, and study materials (including AI chatbot prompts) generated and analyzed during the current study are available in the Open Science Framework repository: [https://osf.io/4g8nx/?view_only=61a1dc34bad144c6a642b2cc11da1815].

References

1. Jolly PM, Kong DT, Kim KY. Social support at work: An integrative review. *J Organ Behav*. 2021;42:229–51. <https://doi.org/10.1002/job.2485>.
2. House JS. *Work stress and social support*. Reading, MA: Addison-Wesley; 1981.

3. Bavik YL, Shaw JD, Wang X-H. Social support: Multidisciplinary review, synthesis, and future agenda. *Acad Manag Ann.* 2020;14:726–58. <https://doi.org/10.5465/annals.2016.0148>.
4. Bankins S, Ocampo AC, Marrone M, Restubog SLD, Woo SE. A multilevel review of artificial intelligence in organizations: Implications for organizational behavior research and practice. *J Organ Behav.* 2024;45:159–82. <https://doi.org/10.1002/job.2735>.
5. Zhang Q, Zhang R, Xiong Y, Sui Y, Tong C, Lin F-H. Generative AI mental health chatbots as therapeutic tools: Systematic review and meta-analysis of their role in reducing mental health issues. *J Med Internet Res.* 2025;27(27i3e78238):e78238. <https://doi.org/10.2196/78238>.
6. Dell'Acqua F, Ayoubi C, Lifshitz H, Sadun R, Mollick ER, Mollick L, et al. The cybernetic teammate: A field experiment on generative AI reshaping teamwork and expertise. *SSRN Electron J.* 2025. <https://doi.org/10.2139/ssrn.5207588>.
7. Pei J, Wang H, Peng Q, Liu S. Saving face: Leveraging artificial intelligence-based negative feedback to enhance employee job performance. *Hum Resour Manage.* 2024;63:775–90. <https://doi.org/10.1002/hrm.22226>.
8. Terblanche N, Molyn J, de Haan E, Nilsson VO. Comparing artificial intelligence and human coaching goal attainment efficacy. *PLoS ONE.* 2022;17:e0270255. <https://doi.org/10.1371/journal.pone.0270255>.
9. Meng J, Zhang R, Qin J, Lee Y-J, Lee Y-C. AI-mediated social support: the prospect of human–AI collaboration. *J Comput Mediat Commun.* 2025;30. <https://doi.org/10.1093/jcmc/zmaf013>.
10. Karinshak E, Liu SX, Park JS, Hancock JT. Working with AI to persuade: Examining a large language model's ability to generate pro-vaccination messages. *Proc ACM Hum Comput Interact.* 2023;7:1–29. <https://doi.org/10.1145/3579592>.
11. Yin Y, Jia N, Waksalak CJ. AI can help people feel heard, but an AI label diminishes this impact. *Proc Natl Acad Sci U S A.* 2024;121:e2319112121. <https://doi.org/10.1073/pnas.2319112121>.
12. Hein I, Cecil J, Lerner E. Acceptance and motivational effect of AI-driven feedback in the workplace: an experimental study with direct replication. *Front Organ Psychol.* 2024;2. <https://doi.org/10.3389/forp.2024.1468907>.
13. Glikson E, Asscher O. AI-mediated apology in a multilingual work context: Implications for perceived authenticity and willingness to forgive. *Comput Hum Behav.* 2023;140:107592. <https://doi.org/10.1016/j.chb.2022.107592>.
14. Sleiman AA, Sigurjonsdottir S, Elnes A, Gage NA, Gravina NE. A Quantitative Review of Performance Feedback in Organizational Settings (1998–2018). *J Organ Behav Manage.* 2020;40:303–32. <https://doi.org/10.1080/01608061.2020.1823300>.
15. Heine SJ, Kitayama S, Lehman DR. Cultural Differences in Self-Evaluation: Japanese Readily Accept Negative Self-Relevant Information. *J Cross Cult Psychol.* 2001;32:434–43. <https://doi.org/10.1177/0022022101032004004>.
16. Wisniewski B, Zierer K, Hattie J. The Power of Feedback Revisited: A Meta-Analysis of Educational Feedback Research. *Front Psychol.* 2019;10:3087. <https://doi.org/10.3389/fpsyg.2019.03087>.
17. Wu S, Zhang Y, Yang H, Tian T. The double-edged sword: investigating when and why supervisor negative feedback promotes versus inhibits employee performance. *J Managerial Psychol.* 2024;39:452–68. <https://doi.org/10.1108/JMP-05-2023-0292>.
18. Watanabe Y, Nakayama M, Takemura K, Uchida Y. AI feedback and workplace social support in enhancing occupational self-efficacy: a randomized controlled trial in Japan. *Sci Rep.* 2025;15:11301. <https://doi.org/10.1038/s41598-025-94985-0>.
19. Watanabe Y, Nakayama M, Takemura K, Uchida Y. Enhancing work engagement through generative AI conversations: a controlled trial on the positive and negative feedback in high vs. low relational mobility environments. *Curr Psychol.* 2025;44:13692–706. <https://doi.org/10.1007/s12144-025-08129-8>.
20. Lee MK, Kusbit D, Metsky E, Dabbish L. Working with machines. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems.* New York, NY, USA: ACM; 2015. <https://doi.org/10.1145/2702123.2702548>.
21. Mori K, Miura K. Development of the Brief Scale of Social Support for Workers and its Validity and Reliability. Volume 9. *Annual Bulletin of Institute of Psychological Studies Showa Women's University*; 2007. pp. 74–88.
22. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. *Psychol Rev.* 1977;84:191–215. <https://doi.org/10.1037//0033-295x.84.2.191>.
23. Kluger AN, DeNisi A. The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychol Bull.* 1996;119:254–84. <https://doi.org/10.1037/0033-2909.119.2.254>.

Figures

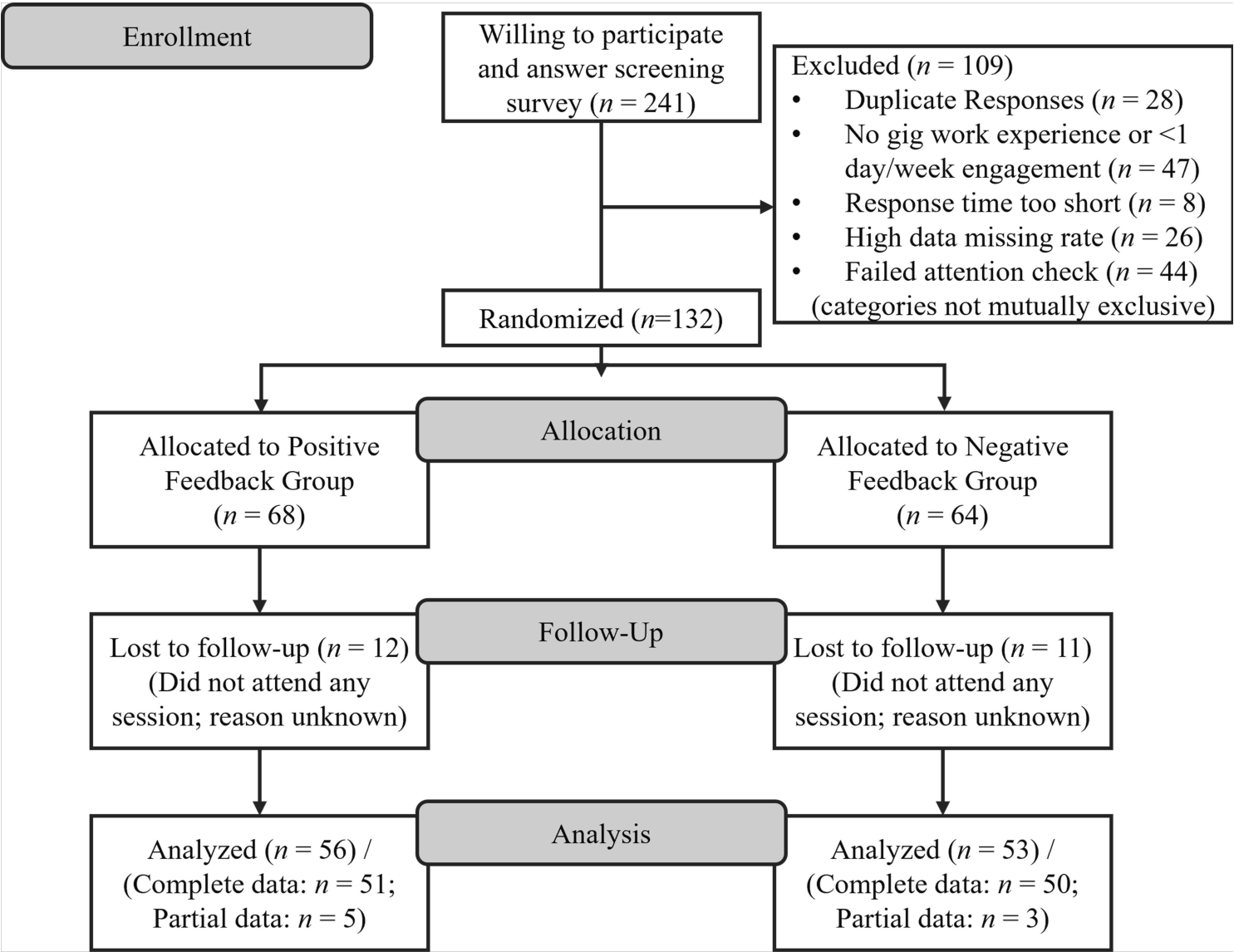


Figure 1

Flowchart of Randomized Controlled Trial

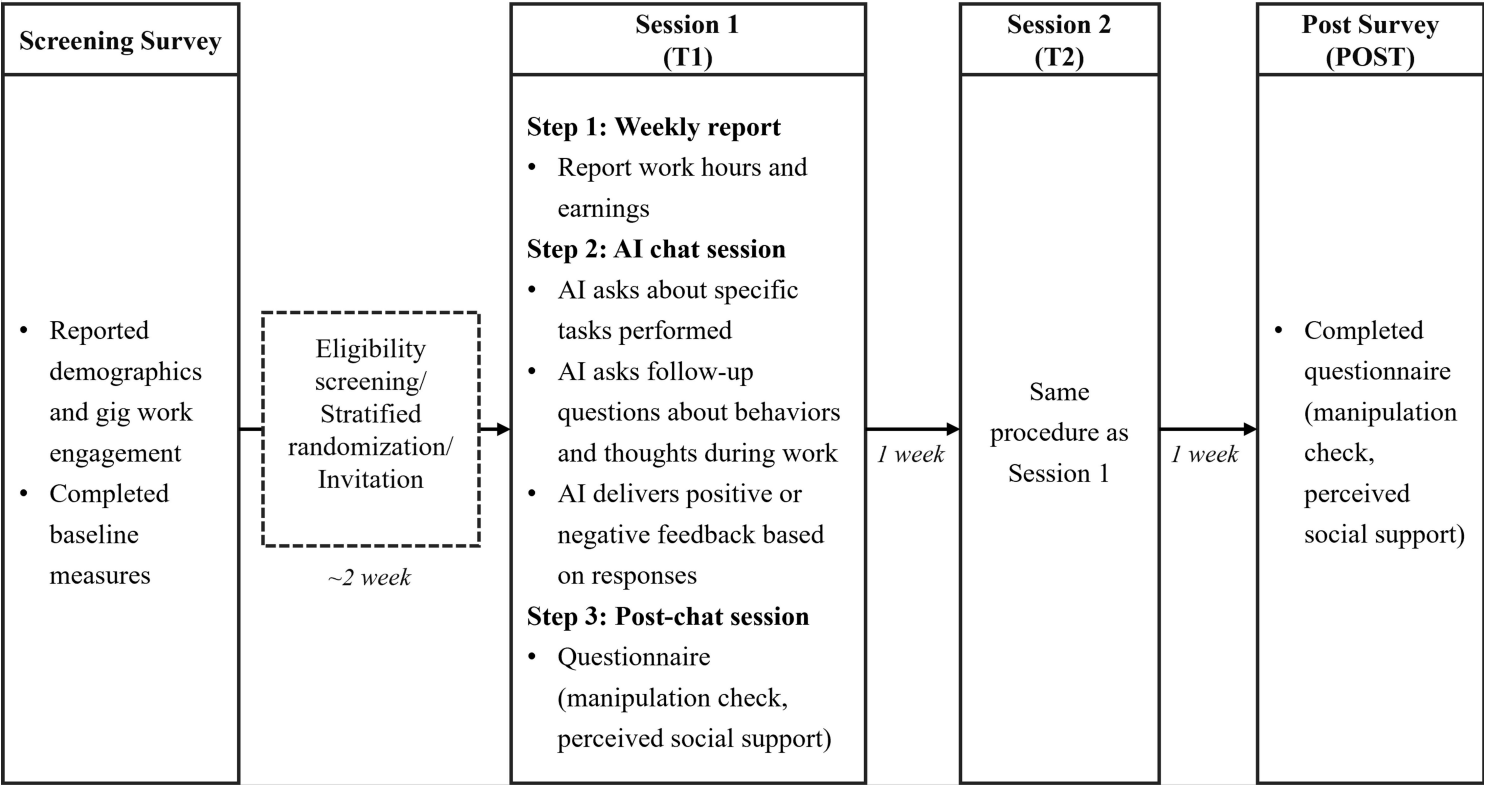


Figure 2

Experimental Procedure and Measurement Schedule

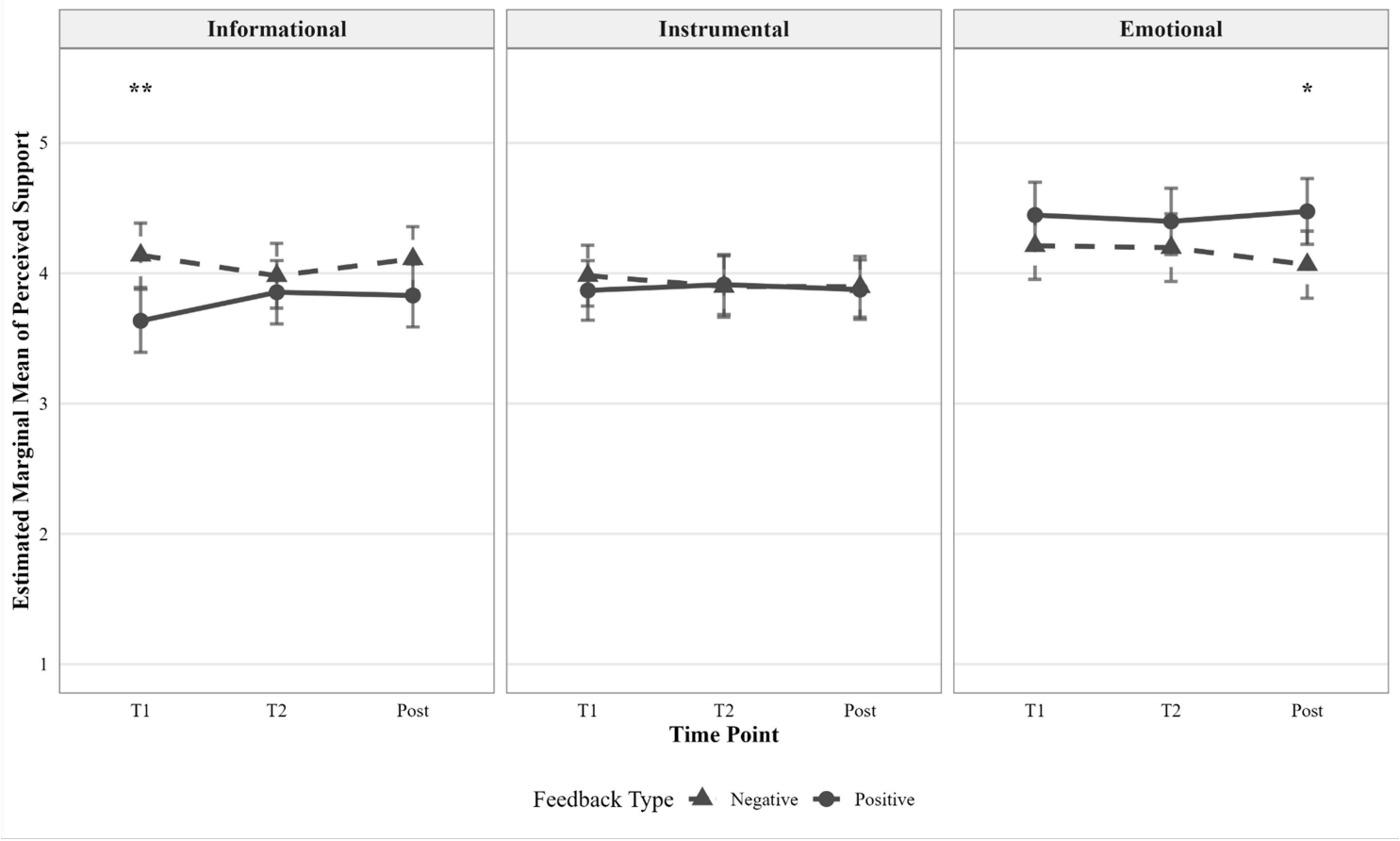


Figure 3

Estimated Marginal Means of Perceived Social Support by Condition and Time Point (Model 2)

Note. Error bars represent 95% confidence intervals. Solid lines with circles indicate the positive feedback condition; dashed lines with triangles indicate the negative feedback condition. Significance markers denote between-condition differences at each time point. * $p < .05$. ** $p < .01$.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [AdditionalfileCONSORT2025checklist.docx](#)