Emotional Load in Service Systems: Definition and Examination of the Effects of Emotional Load on Employee Performance

Research thesis
In Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

Daniel Altman

Submitted to the Senate of the Technion - Israel Institute of Technology
Haifa, March, 2021, Nisan, 5781

The Research Thesis Was Done Under the Supervision of Prof. Anat Rafaeli and Ass. Prof. Galit Yom-Tov in the William Davidson Faculty of Industrial Engineering and Management.

The Generous Financial Help of the Technion the Gutwirth Family, the Crown Family Foundation is Gratefully Acknowledged.

Publications in peer-reviewed journal

Altman, D., Yom-Tov, G. B., Olivares, M., Ashtar, S., & Rafaeli, A. (2020). Do customer emotions affect worker speed? An empirical study of emotional load in online customer contact centers. *Manufacturing & Service Operations Management*. (Study 3)

Refereed papers in conference proceedings

Rafaeli, A., **Altman, D**., & Yom-Tov, G.B. (2019). Cognitive and Emotional Load Influence Response Time of Service Agents: A Large-Scale Analysis of Chat Service Conversations. *Proceedings of the 52nd Hawaii International Conference on System Sciences*. (Study 3)

Conferences

Altman, D., Ashtar, S., Rafaeli, A., & Yom-Tov, G.B. (2017). A Non-Obtrusive Analysis and Evidence of the Costs of Negative Customer Emotion and the Motivating Potential of Positive Customer Emotion. In *Academy of Management.* (Study 3)

Altman, D., Rafaeli, A., & Yom-Tov, G.B. (2017). When Psychology Meets Operations: The Influence of Cognitive and Emotional Loads on the Efficiency of Customer-Service Employees. *INFORMS annual meeting*, Houston, Texas. (Study 3, Finalist at the IBM Service Science Best Student Paper Award Competition)

Altman, D., Yom-Tov, G. B., Rafaeli, A., Ashtar, S., Westphal, M., Natapov., M., & Barkay, N. (2018). New insights on customer sentiment in web-based service interactions. *IOBC*. Tel Aviv, Israel. (Study 3)

My role in the publications listed above was hypotheses and theory development, data wrangling, statistical analysis, and writing.

Contents

List of Tables

List of Figures

List of Equations

Abstract

Workload in service systems is typically examined with limited consideration of differences in expectations, reactions, and demands that customers impose on employees. This thesis proposes that customers vary in the emotional demands they create for employees and that emotional demands influence employee behavior. Studies 1 and 2 examine and define the concept of emotional workload. Until now research used varied definitions for emotional workload, examined a limited set of demands that create emotional workload, and relied on subjective self-reports. I promote a resolution of this ambiguity. Focusing on healthcare work, I show that multiple work events encountered by healthcare employees are emotionally demanding. Next, Study 3 focuses on a different context–contact center employment–and test the effects of emotional workload on service employee behavior. This analysis shows that employees' emotional workload can be objectively estimated.

Study 1 used open-ended interviews with doctors and nurses to identify events that create emotional workload in healthcare work. Interviews identified 260 events that employees reported as creating emotional workload; Study 2 then tested where there is consensus among employees regarding the extent of emotional demand of events. I identified 53 events with such consensus, ranged from creating low to high emotional demand and varied in the frequency with which employees experienced them. For example, expressions of customer aggression cause high emotional demand, while administrative work causes low emotional demand. Work events previously considered as contributing to "workload" or "mental demands" also emerged as emotionally demanding. On the other hand, emotional labor demands, previously considered as the main source of emotional workload, were not uniformly perceived as emotionally demanding. A limitation of Study 2 is that it did not examine the impact of emotional workload on employees, which was the goal of Study 3.

Study 3 tested the effects of emotional workload created by customer negative emotions on the behavior of customer service employees using 141,654 authentic service conversations and Sentiment Analysis software to automatically identify expressions of emotions in text. I found that expressed negative emotion increased employee response time, and positive emotion decreased employee response time. Employee effort—defined as the number of words the employee typed—partly mediated this effect, supporting the argument that emotional demands create workload. The effect of customer expressed emotion was similar in magnitude to the effect of operational workload, suggesting that emotional workload is an important factor to consider when designing workload assignments in service systems. For example, it can be used to improve staffing decisions and to design algorithms that rely on real-time monitoring of emotional workload.

The dissertation expands emotional workload research by demonstrating that it is created by multiple events. The studies identify a range of events that impose emotional demands, with a substantial number of events where employees agree about the level of emotional demand. The findings suggest there are objective elements to emotional workload, challenging current reliance on self-reports. The findings also contribute to affective events theory by showing that employees' emotional workload fluctuates over the workday.

List of Abbreviations and Notations

Abbreviations

OR – operations research
OM – operations management
OB – organizational behavior
RT – response time
Agent RT – the elapsed time between each customer message and the agent response
LOS – length of stay
LIWC – Linguistic Inquiry Word Count
IV – instrumental variable
i.i.d. – independent and identically distributed

Notations

i – index of the customer-agent conversation associated to a case
$NTurns_i$ – the number of turns in conversation i
t – a serial number representing the turn within conversation i
EMO_{it} – the customer emotion expression in turn t, conversation i
RT_{it} – agent response time to a message t in conversation i
δ_i – fixed effect of conversation i
W_{it} – workload related factors that vary during the conversation, captured in turn t, conversation i
u_{it} – error term in turn t, conversation i
$ConvStage_{it}$ – the stage of the conversation, calculated as the turn (t) divided by total number of turns within conversation i
$NumInQueue_{it}$ – a measure of the number of customers waiting in the queue at time t of conversation i
$Concurrent_{it}$ – number of concurrent chats an employee is handling at time t of conversation i
$NumWords_{it}$ – number of words in an agent message sent at time t of conversation i
$\rho_{a(i)}$ – fixed effect of an agent serving conversation i
w_i –error term of conversation i
$CustWords_{1i}$ – the number of words a customer wrote in the first message of conversation i
$IsWeekend_i$ – a binary variable indicating whether conversation i took place on a weekday or a weekend
$HourOfDay_i$ – indicates the hour of the day in which conversation i took place
$SrvType_i$ – service type of conversation i (sales or service)
$ShiftTime_i$ – hours an agent worked when conversation i started
y_{it} – a binary dependent variable which is equal to one if conversation i ends in turn t, and zero otherwise
e_{it} – error term of conversation i at time t for models explaining EMO_{it}
$Turn_{it}$ – ordinal number of current turn t in a conversation i
$CustSent_{it}$ – an additional measure of customer emotion in turn t of conversation i
$\lambda_{d,t}$ – the arrival rate of customers at day d and hour t

1. Introduction

1.1. Emotional Workload in Healthcare: Identifying and Scaling the Emotional Demand in Healthcare Work Events

Healthcare work involves extensive exposure to *emotional demands* that may substantially influence the well-being of healthcare employees, as well as patient safety (Carayon & Alvarado, 2007). Employees experience emotional demands sporadically as a workday unfolds. Furthermore, repeated experiences of emotional demands over time can hamper employee well-being (Felton, 1998), cause emotional exhaustion and burnout (Jackson, Schwab, & Schuler, 1986) and increase nursing staff turnover (Van Der Heijden, Mahoney, & Xu, 2019), a profession that suffers from world-wide shortages. Despite the detrimental impact of emotional demands on healthcare employees, the workload created by emotional demands is poorly defined and insufficiently monitored and managed. In fact, the term "emotional workload" was not mentioned in a recent World Health Organization report titled "State of the World's Nursing" (WHO, 2020).

Reviewing the existing definitions of emotional workload,[1] we found inconsistent definitions, and some of the research claiming to study emotional workload did not even offer an explicit definition (Peräkylä et al., 2015; Rothmann, Mostert, & Strydom, 2006; Voutilainen et al., 2018; Wittels, Johannes, Enne, Kirsch, & Gunga, 2002). Hence, it is currently unclear what circumstances lead to emotional workload or how it can be best measured. One common theme that does emerge from the current research literature, however, is that certain job demands create emotional workload. Job demands are defined as "physical, psychological, social, or organizational aspects of the job that require sustained physical and/or psychological (i.e., cognitive or emotional) effort" (Schaufeli & Bakker, 2004 p. 296). Thus, job demands appear to

[1] The terms "emotional load" and "emotional workload" appear interchangeably in the literature.

be at the core of emotional workload.[2] A second common theme that arises is that emotional workload is typically measured using general self-report questions (e.g., “Does your job demand a lot from you emotionally?”), thus ignoring the possibility of objective emotional demands.

Therefore, the specific job demands that create emotional workload are still uncertain and vary across (the limited) available studies. For example, Drach-Zahavy et al. (2017) referred to the demands placed on employees to manage their emotional expressions at work, a notion that is defined and studied elsewhere as *emotional labor* (Rafaeli & Sutton, 1987, 1989). Carayon and Alvarado (2007) defined emotional workload as “dealing with emotional issues, such as patient death, end-of-life care, and family demands” (p. 122). In the current study, we expand upon these and other previously offered definitions, suggesting that emotional workload can be created by a multitude of demands, including patient aggression (Landau & Bendalak, 2008), patient incivility (Lewis & Malecha, 2011), unrealistic patient expectations (Donabedian, 1988), and abusive supervision (Pradhan & Jena, 2018), as well as role conflict, overload, and ambiguity (Dasgupta, 2012). These types of demands have been previously studied in separate streams of research, however, they have not been examined in connection with emotional workload. Thus, we view the current understanding of emotional workload as limited, and lacking clarity about the job demands that are emotionally demanding and how emotionally demanding specific demands are relative to others. It is also unclear whether specific emotional demands are experienced similarly by different employees. Answering these questions is our goal in Studies 1 and 2. We rely on affective events theory (AET; Weiss & Cropanzano, 1996)–an event-based approach to the study of emotions at work–to examine emotional workload in healthcare work. AET proclaims that employees' affective experiences change in relation to specific work events. However, AET research has

[2] I refer to job demands that create emotional workload as “emotional demands” or “emotional job demands” interchangeably.

focused primarily on employees' positive and negative affect after experiencing a specific event (c.f., Ohly & Schmitt, 2015). In contrast, we argue that some events are emotionally demanding and, therefore, create the additional affective experience of emotional workload.

Building on the granular event-based perspective of AET, we propose a broad definition of emotional workload that (a) includes *all* emotional demands, (b) acknowledges that different demands can differ in the level of emotional demand they impose, and (c) frames emotional workload as a construct that can be measured objectively. Attending to the call of Brief and George (1995) to identify experiences that are common among employees, we promote a way to objectively assess emotional workload. We draw upon the operations research (OR) definition of operational workload, as the number of customers multiplied by the amount of work each customer requires (Hall, 1991). Applying this definition to emotional workload, we offer a framework that parallels the count of customers to the count of events that create emotional demands and the amount of work that each customer requires to the level of emotional demand that each event poses. In other words, we define emotional workload as *a function of the number of emotional demands and their level of demand.* Our analysis promotes the implementation of such explicit definitions by identifying events that are deemed to be emotionally demanding by healthcare workers (Study 1) and by estimating the level of demand that each event creates (Study 2).

Study 1 is a qualitative study, which identified 260 events representing job demands that may create emotional workload for healthcare employees. Study 2 used crowdsourcing to collect data from healthcare employees to assess the extent to which each of the Study 1 events is emotionally demanding, to identify events that employees agree upon the level of emotional demand they create, and to assess the frequency with which healthcare employees experience the different demands.

The contribution of our analyses in Studies 1 and 2 is threefold:

1) We deepen the understanding of healthcare work by identifying events that are emotionally demanding and showing that some operational demands can also be emotionally demanding. As such, we argue that workload management needs to account for emotional demands above and beyond operational parameters.
2) We expand the understanding of emotional workload by identifying events for which there is consensus among employees regarding the level of emotional demand. We view such consensus as identifying the “objective” or shared component of emotional workload. Thus, we expand on past research that considered emotional workload to only be a subjective experience. This novel approach to emotional workload can allow researchers and managers to objectively monitor emotional demands.
3) We extend AET research beyond the focus on positive and negative affect created by events by demonstrating the large array of events that can create an experience of emotional workload in healthcare work.

1.1.1 **Theoretical Framework**

Employees can experience various levels of emotional workload as they encounter different job demands. Such variations have not been considered in past research. Identifying and coding job demands that are emotionally demanding may have been avoided in past research because it is costly. Consequently, the common approach has been to obtain *overall appraisals of* emotional demands. Past studies have asked employees to think about their work in general (Bakker, Demerouti, & Schaufeli, 2003) or over the past day (c.f., Donoso, Demerouti, Garrosa Hernández, Moreno-Jiménez, & Carmona Cobo, 2015). Researchers then ask participants to respond to questions such as, “Does your work demand a lot from you emotionally?” on a scale of 1 (“not at all”) to 7 (“to a great extent”) (cf. Llorens, Bakker, Schaufeli, & Salanova, 2006). This approach allows easy access to employees' perceptions of emotional demands without identifying specific events that create these demands. Furthermore, it does not connect the emotional demands to a specific work context. However, there are several theoretical and practical disadvantages to this approach.

First, a single overall assessment assumes that different emotional demands have a similar impact on employees. Yet, as Crawford et al. have stated (2010), some demands positively impact employees ("challenging demands"), whereas others have a negative impact ("hindering demands"). Moreover, as Crawford et al. (2010) noted, "[emotional demands are]…difficult to classify as either challenges or hindrances…" (p. 838). Hence, we propose that a focus be placed on specific demands and the identification of the emotional workload created by each demand; such an approach would allow for bypassing the similar effect assumption.

Second, self-report appraisals are subjective and known to be influenced by multiple biases. For example, an employee who experiences a very demanding event just before responding to a survey is likely to be influenced by the availability of this event in his or her memory (Tversky & Kahneman, 1973). Thus, self-report responses could be confounded by external aspects that are irrelevant to an employee's experience. However, some demands may be experienced similarly by different employees and have a similar impact on them (Brief & George, 1995). Identifying such (consistent-impact) demands can provide a more objective measure of emotional workload.

Third, measuring emotional demands through self-reports is highly obtrusive and, therefore, provides a low-resolution view of a single and arbitrary point in time (Bakker et al., 2003) or, at best, daily assessments conducted over a few days (Donoso et al., 2015). Such measures cannot capture the cumulative effects of one's encounters with multiple sporadic emotional demands over a workday. We promote a measurement system that embraces AET, thereby developing an event-based analysis of emotional demands in healthcare work.

1.1.2 **Affective Events Theory**

AET provides an overarching framework for studying emotion at work by considering that emotionally charged events can lead to affective changes in employees. Researchers have built on AET to study workplace incivility (Schilpzand, De Pater, & Erez, 2016) and workplace aggression (Rafaeli et al., 2012), among other topics. AET proclaims that the multiple events employees encounter at work drive changes in their emotional responses (Weiss & Beal, 2005). AET labels these responses *emotional proximities* of events, arguing that such proximities have an immediate and time-bound impact on employees. Until now, only positive and negative affects resulting from specific events have been studied as emotional proximities (see Ohly & Venz, 2021 for a summary). To the best of our knowledge, previous research has not considered the broader notion that work events occurring during healthcare employees' workdays can create emotional demands, consequently adding to their workload.

That said, we highlight two studies that identified specific events that immediately affected employees' sense of fatigue and effort and, therefore, may be viewed as contributing to emotional workload. Zohar, Epstein, and Tzischinski (2003) showed that goal-disruptive events (e.g., events that disrupt one's scheduled activity) were related to medical residents' immediate sense of fatigue. They also found that some goal-enhancing events (e.g., encountering a medically interesting issue) were associated with residents' fatigue when the operational workload was high. Relatedly, in Study 3 we report that expressions of customer emotion in a text-based service conversation influenced employee efforts and response time. These two studies illustrate our point that specific work events can be emotionally demanding and, thus, can influence an employee's workload. However, these studies identified only a few of the myriad of events that may be emotionally demanding for employees. Additional events that create emotional workload are likely to exist, and the extent to which each event is emotionally demanding remains unknown. These gaps led us to the following research questions:

Q1: What events in healthcare work pose emotional demands?

Q2: What is the relative level of emotional demand for each of these events?

2. Study 1 - Identifying Emotionally Demanding Events in Healthcare Work

2.1. Method

Using the critical incident technique (Butterfield, Borgen, Amundson, & Maglio, 2005), we conducted semi-structured in-depth interviews with healthcare employees between September 2019 and July 2020. We recruited participants through Facebook groups of nurses and doctors, as well as via snowball sampling (Penrod, Preston, Cain, & Starks, 2003). We continued interviewing until reaching theoretical saturation of the data (Guest, Bunce, & Johnson, 2006), which resulted in 24 interviews (79.2% women, *n*=19). Twenty participants worked in hospitals and 4 worked in HMOs [M*age*=37.08 (SD=10.51), M*tenure*=9 years (SD=9.1)]. Half of the participants were nurses, and the other half were doctors. Interviews ranged from 26 to 62 minutes (*M*=39.17 minutes, *SD*=8.93); the 24 interviews together totaled 15 hours and 40 minutes. Participants received a coffee voucher at the beginning of the interview and then signed a consent form, which included a request to record the interview. Participants were informed that they could withdraw from the study at any point and could request that all or some of their interview not be used in the study. Two participants asked us not to record a small part of their interview and two other participants asked us not to record the interview at all. However, all participants allowed us to include all of the interview content in the study.

Interviews began with questions that aimed to prompt participants to think about their general workload. The three broad questions were as follows:

(1) Please indicate events that create workload at your work.

(2) Are there different types of workload? Can you provide examples of each type of workload?

(3) How do these types of workloads impact your work?

We allowed interviewees to elaborate and focus on what they viewed as relevant or important in their responses, and then we probed further with the following questions:

(4) Can you describe a work-related event that "stuck" with you for a while (i.e., you could not "let go of it")?

(5) Can you describe an event that you think interrupted your professional work or the work of a colleague?

(6) Can you describe an event where you felt threatened?

Before ending the interview, we debriefed participants about the goal of the study and asked for any further comments:

(7) Our goal in this study is to identify events that cause emotional workload. Are there any other events that you can think of which you did not mention and might be relevant?

We ended the interview with a positive question to defuse any tension that the discussion may have caused for participants:

(8) Can you tell me what you like most about your job?

2.2. **Results**

Twelve interviewees mentioned emotional demands as creating "workload" in their responses to the first question and 17 (70.8%) explicitly mentioned "emotional workload" as a type of workload in the second question. In total, the interviews yielded a list of 260 distinct events mentioned by respondents as creating workload in their job (see Appendix 1 for the full list of events, by categories).

Some of the events noted appeared to be similar to those mentioned in previous research on emotional workload. For example, emotional labor events (e.g., being unable to express an angry feeling), death events (e.g., providing treatment to a patient who is about to die), and patients' family issues (e.g., being threatened by a patient's family member). The majority of the identified events, however, were not mentioned in past research as emotionally demanding. For example, legal-related events (e.g., sending a patient for a test only to avoid a possible lawsuit), professional challenges (e.g., providing a treatment without proper experience), co-worker issues (e.g., hearing other staff members speak in an unfamiliar language), and system issues (e.g., witnessing another department receiving high scores for patient satisfaction).

Importantly, the events identified in the interviews were all referred to as some form of workload (i.e., emotional or operational). However, some events may create an operational workload that is not emotionally demanding (i.e., having to perform bureaucratic work) and other events may create high emotional demand for some employees but not others. Thus, our next goal was to identify events that created emotional workload for multiple employees, with the following research question:

Q3: What events are agreed upon by employees regarding their level of demand?

Study 2 used crowdsourcing to collect data from a larger sample of healthcare employees to estimate (a) the extent to which events mentioned in Study 1 are emotionally demanding, (b) the extent of agreement of healthcare employees regarding the level of emotional demand of events, and (c) the frequency with which events are experienced by healthcare employees.

3. Study 2 - The Extent of Emotional Demand of Work Events in Healthcare Work

Motowidlo, Packard, and Manning (1986) identified 45 events that correlated with nurses' overall stress. For example, they found that the frequency of experiencing an event in which "a doctor wastes your time by having you perform non-nursing tasks" (p. 629) was correlated (r=0.3) with a self-report stress index. Although such an event can be emotionally demanding, a significant correlation between the frequency of experiencing an event and overall self-reported stress does not identify the extent to which specific events are emotionally demanding. We note three methodological issues in this study.

First, self-report stress measures refer to employees' overall job experience, hence evaluating the stress from some undefined aggregation of experiences without distinguishing between specific events. Second, Motowidlo et al. (1986) reported the correlations between two self-report measures to identify stressful events, a method likely inflated by common method bias (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). Third, the significance of a correlation depends on the variance in the experience of a specific event. If a certain event rarely occurs, a low variance would yield a low correlation, even though the event could create a high emotional demand when experienced. For example, a case in which a patient threatens a nurse with a weapon, although rare, clearly imposes high emotional demand. Study 2 overcomes these limitations by collecting assessments of the emotional demand of each specific event identified in Study 1 from multiple raters.

3.1. **Method**

Participants

We recruited 126 healthcare employees using Prolific (https://www.prolific.co/). The selection criteria included current employment in healthcare work and English fluency. The sample comprised 78.2% women, 55.6% nurses, 14.5% doctors, 8.9% medical support staff, 7.3%

administration workers, and 13.7% other types of healthcare workers, including radiographers, therapists, and technicians. Participants were working in hospitals (69.4%), HMO's (8.9%), clinics and general practices (8.9%), mental health centers (6.5%), and other organizations such as care homes and hospices (6.5%). Most participants had tenure of over ten years (60.5%), and the remainder had tenure of 7-9 years (9.7%), 4-6 years (18.5%), or three years or less (5.6%). Participants' mean age was 41.39 (SD=11.96). Participants were rewarded with £1.25 for their participation in the study, and spent an average of 8 minutes completing the study.

Procedure and Tools

After signing a consent form, participants were told, "Imagine a typical workday at your workplace. Suddenly, the following situation occurs…" This statement was followed by an event selected randomly from the 260 events identified in Study 1. Participants were then asked to rate (1) the extent to which the event would be emotionally demanding to them and (2) the frequency with which they experience such an event in their work. Each participant rated ten events. Demographic information was collected after the ratings.

Emotional Demand

We adapted six items used in previous studies of emotional demands (Van Veldhoven & Meijman, 1994), from assessing individual-level demand ("Does your work demand a lot from you emotionally?"), to assessing event-level demand ("Does this situation demand a lot from you emotionally?"). Responses ranged from 1 ("not at all") to 7 ("to a great extent"). Cronbach's Alpha was 0.89. The emotional demand score is the average of all items.

We measured the frequency with which the participants experienced the event with the following question: "How often do you experience such a situation in your work?" Response options were “never”, “once or twice a year”, once or twice a month”, “once or twice a week”, “three or four times a week”, “almost every day”, and “multiple times a day”.

3.2. **Results**

The goal of Study 2 was to identify events that are commonly perceived as emotionally demanding and, therefore, can be assumed to create emotional workload for the general population of healthcare employees. We asked participants to rate the 260 events reported in the Study 1 interviews and obtained a total of 1,181 ratings for all of the events. We eliminated ratings in which raters noted that they had never experienced the event (396); this resulted in 31 events that had fewer than 3 ratings (the minimum that we required) which were also removed. The final sample included a total of 736 valid ratings of 229 events in which at least 3 different raters reported on each event (M_{raters}=3.21, SD_{raters}=0.41). The emotional demand of the events was rated as high (33.89%; ≥ a score of 5), medium (43.22%; score between 3 and 5) or low (22.89%; a score ≤ 3). For the sake of brevity, we report of events experienced with “high” frequency (more than once a month) and events experienced at “low” frequency (less than once a month). Participants reported experiencing 394 events with high frequency, and 391 events with low frequency.

Inter-rater agreement among ratings of the same event reflects the extent of consensus between participants regarding the emotional demand that an event creates (James, Demaree, & Wolf, 1984). Therefore, we screened for events with inter-rater agreement of $r_{wg(j)}$=0.7 or higher, which yielded 53 events. We construe these events as events for which healthcare employees generally agree about the emotional demand that they create. Table 1 lists these 53 events, their experienced frequency, and their emotional demand rating. Figure 1 summarizes the frequency with which employees experienced these 53 events for each level of emotional demand (low/medium/high). The figure illustrates that there are events that all raters experienced with either high frequency (25%) or low frequency (15%); on the other hand, the majority of events were experienced with mixed frequency (60%). Noteworthy is that, all frequency categories were present across all three levels of emotional demand. In addition, of the low emotionally demading events (12 events), 11 were reported to be highly frequent by at least some raters. Of the high emotionally demanding events (21 events), only one event was indicated as highly frequent by all raters, but 16 events were rated as highly frequent by at least one rater.

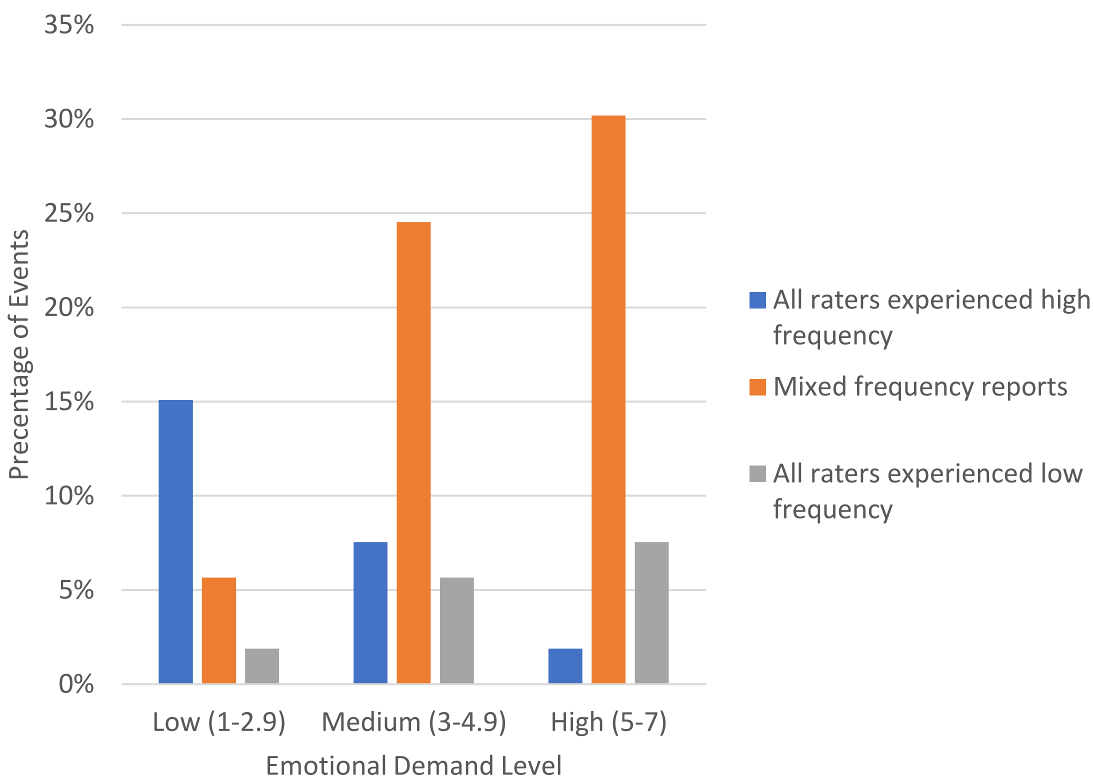


Figure 1 - Distribution of events with low, medium, and high emotional demand by the frequency that they are experienced by healthcare employees.

Interestingly, raters did not agree on two subsets of events in regard to the emotional demands they pose. First, there was no consensus about the emotional demand of any "emotional labor" events. For example, being unable to express anger received ratings of 2.66, 4.33, and 3.83 ($r_{wg(j)}$=0). This challenges the presumption in past research that emotional labor is a primary cause of emotional workload (c.f., Drach-Zahavy et al., 2017). Second, there was no consensus about the emotional demand ratings for the "error" events. For example, making a mistake in diagnosing a medical condition was given ratings of 3.00, 4.71, and 5.14 ($r_{wg(j)}$=0).

In contrast, we found that events not previously considered to create emotional workload were rated as emotionally demanding. For example, multitasking, previously considered as a source of cognitive load (Czerwinski, Horvitz, & Wilhite, 2004) or "mental demands" (Bakker & Demerouti, 2007), emerged in our data as being moderately emotionally demanding. For example, "You are assigned multiple new patients at the same time" (3.72/7.0) and "You need to reprioritize your tasks" (4.33/7.0) were rated as 3.72 and 4.33, respectively.

Table 1 - Healthcare work events, experienced frequency, and ratings of emotional demand

Event	Freq.	Emotional Demand
A patient physically attacks you	Low	6.89
A patient approaches your work station aggressively	Mixed	6.63
A doctor expresses distrust towards you	Low	6.56
A patient's family member is threatening you	Mixed	6.50
A patient's family member is yelling at you	Mixed	6.42
You feel like you are being blamed for problems you cannot solve	Mixed	6.39
Your manager tries to change your working conditions without your consent	Mixed	6.33
A patient physically threatens you	Mixed	6.29
You tie a patient to the bed to prevent self-harm	Mixed	6.22
Security arrives because of a patient's aggressive behavior	Mixed	6.22
Your patient implicitly threatens to sue you	Low	6.17
Your patient complains about you in front of other people	Low	6.00
You have to work without proper equipment for protection from COVID-19	High	5.78
A staff member undermines you in front of patients	Mixed	5.72
Your manager does not back you up	Mixed	5.72
You suspect that your child patient is being abused by a parent	Mixed	5.67
You must deal with a dissatisfied patient	Mixed	5.46
You send someone for a test just because you are afraid of a lawsuit	Mixed	5.38
A patient's family member enters your break/lunch room	Mixed	5.33
Family expectations increase because a patient who is about to die suddenly improves but you know it is only temporary	Mixed	5.17
A staff member doesn't arrive at all to the shift where you were supposed to work together	Mixed	5.06
A patient refuses treatment from you specifically	High	4.83
Your patient requests something immediately, and it is not possible	Mixed	4.83
Your manager is unfair to you	Mixed	4.72
The staff in your department work inefficiently	High	4.56

You need to reprioritize your tasks	High	4.33
A patient refuses to be treated by you because of your gender	Low	4.22
You deliver bad news to a patient	Mixed	4.22
You provided a treatment you are inexperienced in giving	Low	4.17
You got an urgent request to come to work	Mixed	4.11
You suspect the medical diagnosis of a patient but cannot tell the patient until more tests are run	Mixed	3.92
A staff member is calling your personal phone during work and you must answer it	Mixed	3.89
Your shift has ended but you must continue working	Mixed	3.78
You are assigned multiple new patients at the same time	Mixed	3.72
A staff member doesn't arrive on time for their shift where you are working together	Low	3.67
You provide treatment that will likely harm a patient who will die soon anyway	Mixed	3.61
There is a miscommunication between you and another staff member	High	3.56
A patient enters a treatment room where you are treating another patient	Mixed	3.50
A staff member intervenes in your tasks	Mixed	3.44
You have multiple managers	Mixed	3.33
A patient moved to the front of the queue because of his/her medical condition	Mixed	3.28
There are patients "chattering" outside the treatment room you are working in	Mixed	2.83
You and another staff member are having a private "venting" conversation	High	2.72
You are waiting to consult with a senior doctor	High	2.50
You were assigned a new patient	High	2.39
Another department receives higher scores for patient satisfaction	Low	2.39
Your tasks are distributed between distant locations	High	1.79
Someone knocks on your door while you are with a patient	High	1.67
You call a doctor who is in a different ward	Mixed	1.61
Other staff members speak in a language you don't understand	Mixed	1.56
You have to perform bureaucratic work	High	1.50
You check your email repeatedly	High	1.33
A patient with COVID-19 asks you to bring her water because she is alone in quarantine	High	1.28

3.3. **Discussion**

Using both qualitative and quantitative methods, we identified 53 real-life events about which healthcare employees agree regarding the extent of emotional demand that they pose (Table 1). We posit that these events can be assumed to create emotional workload when a healthcare employee experiences them. Some of these events posed *low* levels of emotional demands (e.g., performing bureaucratic work; an average score of 1.5), others *high* levels of emotional demands (e.g., being yelled at by a patient's family member; average score of 6.42), still others *moderate* levels of emotional demand (e.g., providing a treatment with which you are inexperienced; average score of 4.17). The frequencies that these events occurred also vary.

The set of events that we identified as creating emotional workload is broader than the set reported in previous research (Carayon & Alvarado, 2007), which focused on family demands and death. We showed, for example, that demands previously classified as "mental workload" (e.g., multitasking; Bakker & Demerouti, 2007) also pose emotional demand. Thus, both future research and the management of emotional workload must account for a multitude of issues that have not been considered until now. Importantly, previous research has studied many of the events we identified. However, we offer a conceptual and empirical framework of emotional workload as a way to unify and standardize the examination of the different events. We also offer empirical data on the relative demand that each event poses. This approach allows us to connect and compare different phenomena, such as patient aggression and professional challenges.

The lack of agreement regarding the emotional demand of emotional labor events is surprising given the assumption in the literature that emotional labor equates to emotional workload (Drach-Zahavy et al., 2017). This finding suggests that only some people experience emotional labor as emotionally demanding, and is consistent with Donoso et al.'s (2015) claim that the impact of emotional labor on employees depends on individual differences in emotion regulation abilities. Thus, our analysis reinforces the need to consider emotional regulation capabilities and, perhaps, other individual differences in both the analysis of emotional labor specifically and, more broadly, in the study of emotional workload. Furthermore, the lack of consensus regarding error events could be explained by Russo, Buonocore and Ferrara's (2015) finding that reporting an error is emotionally difficult; thus, employees may avoid admitting to making a mistake, even to themselves. This finding again challenges the use of self-report measures for the study of emotional workload and might suggest that people may avoid acknowledging emotional demands in order to regulate their own sense of emotional workload.

Our studies extend AET research to suggest that emotional workload is an affective experience that changes according to the events that employees encounter at work. Similar to fluctuations in employees' positive and negative affect identified in previous AET research (see Ohly & Venz, 2021 for a summary), we show that emotional workload can also fluctuate as employees' workdays unfold. Therefore, future research on emotional workload must recognize and embrace the temporal aspect of emotional workload among employees.

3.3.1 **Toward More Objective Measurement of Emotional Workload in Healthcare**

Our analyses distinguish between events where employees are in agreement regarding the level of demand and events that employees experience differently. We see this as a step toward an objective view of work experiences, as called for by Brief and George (1995), and posit that events with high agreement represent "objective" instances of emotional demand. The ability to objectively identify and categorize events is useful for both management and research because it lays a foundation for automated estimation of the emotional demands an employee experiences.

In Study 3, we use automated analyses to identify and code expressions of customer emotion to assess the emotional workload of service agents. Similar to records of customer service texts, some "high-agreement" events are typically recorded in hospital information systems. As illustrated in Figure 2, the emotional demands a healthcare employee might encounter over a workday can be identified in relation to the events experienced. Thus, each event's level of emotional demand can be used to estimate the employee's total workload in a given time. However, this is not a simple calculation and assuming these demands to be additive is likely too simplistic. It could be that events with very high emotional demand outweigh events with lower emotional demands (Miron-Shatz, 2009), and that the effect of an event may depend on when it occurs (e.g., time of day, point in shift, whether experienced simultaneously with other events).

3.3.2 Managerial Implications

Measuring employees' emotional workload can be managerially useful. In measuring the different events experienced by the employees, our analyses illustrated the possibility of relying on technological means to objectively identify and code emotional demands, thereby tracking

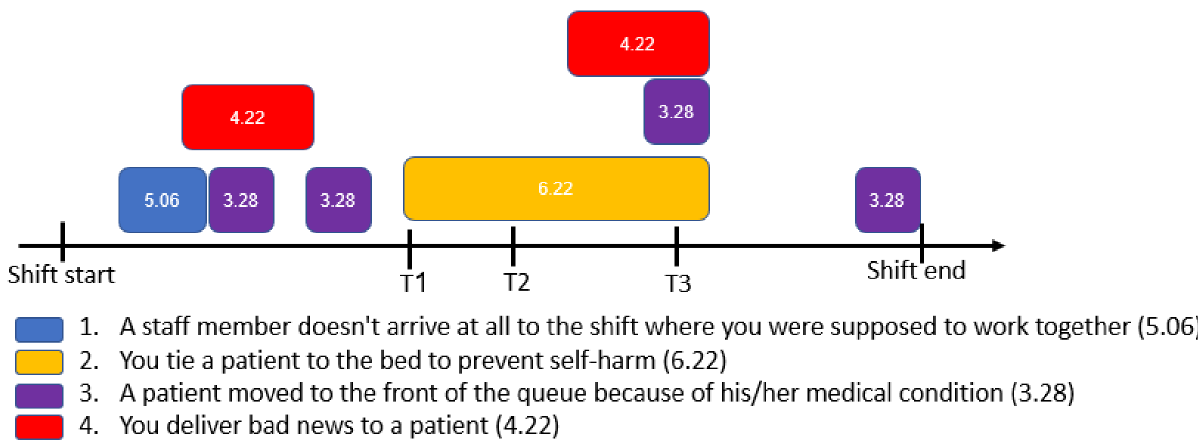


Figure 2 - Illustrating one employee's emotional demands during a workday. Colors represent events and numbers indicate the level of emotional demand of each event. Rectangles start when an event occurs, and end when the event no longer poses emotional demand.

employees' emotional workload. Computer records already include some of the emotional demands that we identified and can provide insight into employees' workdays (e.g., tying a patient to bed to prevent self-harm). In addition, the use of computer vision and voice analysis (c.f., Kooij, Liem, Krijnders, Andringa, & Gavrila, 2016) can allow for the automatic identification of employees who experienced aggressive behaviors of patients and their family members using face recognition software (e.g., "Amazon Rekognition"; Amazon Web Services, 2021). Such tools would allow for automatic and continuous measures of emotional demands, creating an objective measurement of emotional workload. By implementing novel managerial tools, employees who experience an unusual amount of emotional workload can be flagged, as well as departments that need further managerial attention and emotional support. Applying this approach to healthcare could promote a supportive and mindful environment, which addresses emotional workload, facilitates employee recovery and reduces burnout, thereby increasing operational efficiency.

3.3.3 Limitations and Future Research

A question that we were not able to examine in Studies 1 and 2 was the duration of the emotional demand effects of events, which is important for analyzing cumulative emotional workload. We note that the timeframe of the effect may vary across events or people and recommend that this be examined in future research. Further research is also needed to assess the effects of various employee resources on the accumulation of emotional workload (c.f., Donoso et al., 2015) or, relatedly, of strategies that employees can utilize to accrue resources (e.g., Fritz, Lam, & Spreitzer, 2011) to cope with their emotional workload. We note that available studies of employee resources (similar to emotional demands) tend to examine *perceptions* rather than actual resources (cf., Bakker & Demerouti, 2007) and that they reflect specific and arbitrary points in time, thus, not addressing potential variations over time. Hence, further research on the dynamics of employee resources is needed to connect employee resources to emotional workload. A third limitation is that the different cultural backgrounds of employees may have influenced their experience of the same emotional demands, as they interpreted them through the lens of their unique cultural and social norms, and may have caused them to react differently (De Vaus, Hornsey, Kuppens, & Bastian, 2018). Such cultural differences were not considered in the current research. Finally, Studies 1 and 2 did not examine the impact of actual experiences of emotional demands on employees or their work performance. This is the Focus of Study 3.

4. Study 3 - Do Customer Emotions Affect Agent Speed? An Empirical Study of Emotional Workload in Online Customer Contact Centers (Study 3)

4.1. Literature Review

Toward understanding agent efficiency in service systems, research in Operations Management (OM) has investigated the impact of operational workload on agent efficiency (Kc & Terwiesch, 2009; Song, Tucker, & Murrell, 2015). Studies on the effects of operational workload are inconclusive, with some research indicating it increases efficiency and others showing it decreases efficiency (see Delasay, Ingolfsson, Kolfal, & Schultz, 2019 for a review of mechanisms that might explain this confusion). In the spirit of incorporating human behavioral aspects into OM research (Cho, Bretthauer, Cattani, & Mills, 2019), we propose that the behavior of agents and emotions that customers express, generally ignored in operations research (Field et al., 2018), should be added to this discussion. We suggest that this salient aspect–of customer expressed emotion–can promote the understanding of agent performance-related behaviors (e.g., speed, effort) and help improve understanding and management of service delivery.

Research in Organizational Behavior (OB) describes the effects of emotions that people express toward other people, be it in negotiations (van Kleef, De Dreu, & Manstead, 2004), or in other forms of social interactions (Hareli & Rafaeli, 2008). Lab experiments, for example, show that customer emotions affect the speed, accuracy and fatigue of service agents (Rafaeli et al., 2012). Studies also show that negative customer emotions lead to agent incivility (Walker, van Jaarsveld, & Skarlicki, 2017), and that the amount and valence of emotions that customers express influence service agents (Grandey, Dickter, & Sin, 2004; Grandey, Rafaeli, Ravid, Wirtz, & Steiner, 2010). Building on such research and the results presented in Studies 1 and 2, we conceptualize emotional workload as the amount of emotional demands that a service agent encounters and must handle. Emotional workload complements the construct of operational (or offered) load, recognizing and incorporating behavioral variability between people into analyses of service work. Emotional workload adds an additional dimension to the load that service delivery agents experience.

This research offers five contributions to current research on service delivery. First, we provide evidence additional to Studies 1 and 2 that emotional workload can be estimated by coding specific emotional demands. Namely, emotions that customers express to agents. Second, we show the effects of emotional workload on operational measures, notably agent response time to customers and number of turns a service interaction requires. We show these effects are above and beyond the effects of operational load. Third, we investigate one of the mechanisms that explains the effects of emotional workload, agent effort. Fourth, we examine both the influence of customer emotions on service agents' behavior (i.e., response time) and the subsequent influence of that agent behavior (i.e., response time) on customer emotions, within the same data. Finally, we use automated sentiment analysis to analyze customer emotions in a large sample of authentic service conversations. Our analyses provide important foundations for evaluating efficiency and optimizing work allocation. To the best of our knowledge, this is the first study to recognize and analyze the dynamic nature of the emotionally-charged customer-agent conversations.

Our paradigm overcomes multiple biases and limitations of previous research (Donaldson & Grant-vallone, 2002), which was conducted primarily by OB scholars, and relied extensively on lab simulations (cf., Rafaeli et al., 2012; van Kleef et al., 2004), and self-report measures (cf., Wang, Liao, Zhan, & Shi, 2011). By using automated sentiment analysis (Thelwall, 2013; Yom-Tov et al., 2018), we obtain unbiased measures of customer emotion from real-life data, and provide clear operational and managerial implications. We analyze individual messages within customer-agent conversations as instances of customer expression of emotion and agent work behavior. This focus offers high resolution into the dynamics within conversations. Also, our findings expand beyond the impact of negative customer emotions, which has been the focus of most research, to include also the effects of positive customer emotion (Goes, Ilk, Lin, & Zhao, 2018).

The context of our study is contact-center service, which is technology-mediated, and allows access to detailed data and measures of both agent and customer behavior (Rafaeli et al., 2017; Rafaeli, Ashtar, & Altman, 2019). Specifically, we analyze 141,654 customer-agent conversations from the archives of a large western transportation company. We empirically test the impact of emotional workload created by customers on (a) agent response time to customers, (b) agent effort, and (c) number of turns/iterations required to complete the service.

Our main dependent variable is agent response time (RT) to a specific message of a focal customer. A key challenge we embrace is estimation of causal effects using the variation within service conversations. Our analyses show that higher emotional workload, in the form of negative customer emotion, increases agent RT and the effect is 2.66 times larger in magnitude than the effect of agent multitasking, and stronger than system-level load (queue length). Negative customer emotion increases the length of text in agent replies by 4.3% and positive emotion increases the length of text in agent replies by 2%, compared to the text of neutral message. In addition, a one-point increase in negative customer emotion increases agent RT by 19.7%. Considering the reverse effect of agent RT on customer emotion, we show that if the agent doubles the RT, customer emotion decreases by about 0.1 standard deviations. This finding has implications for acceptable levels of in-service waits (i.e., waiting during customer length of stay) that result from concurrency decisions.

4.1.1 **Context of the Study and Data Description**

The current study is based on data provided by LivePerson Inc., a firm that offers a web-based service platform. The platform allows end customers to interact with agents of a service brand, through written “chat” messages. Customers who want to chat with a live agent enter a queue and wait for an available agent. Service chats comprise iterations of agent and customer written messages.

A feature unique to chat service platforms is that agents can simultaneously interact with multiple customers (maximum of 3 customers in our data). Agents waiting for a focal customer to respond can turn to interact with other customers. The implication is that if an agent is busy with one customer, his or her other concurrent customers must wait. Customers are not explicitly informed of this agent multitasking and do not know why an agent’s response is delayed.

4.1.2 Data Description and Definitions

Our data includes 141,654 service conversations conducted from March 2016 to April 2017, by agents of a western transportation company. We use the terms “chat” and “conversation” interchangeably to refer to a full service interaction between an agent and a customer. Each conversation in the data includes agent and customer lines, as well as system lines, which are automatically generated, and not included in our analyses since they do not reflect any human input. The term “line” refers to a single parcel of text sent by a customer/agent (i.e., followed by pressing “enter”) and “message” refers to one or more lines sent, uninterrupted, by a customer or agent. That is, a series of lines sent by an agent or customer are collapsed into one message. Figure 3 offers a schematic view of the simultaneous chats of one agent who is handling three customers, where each chat comprises multiple messages. Chats in our sample last on average 11.7 minutes (SD=9.46), and include on average 5.40 customer messages (SD=3.54) and 5.78 agent messages (SD=3.50).

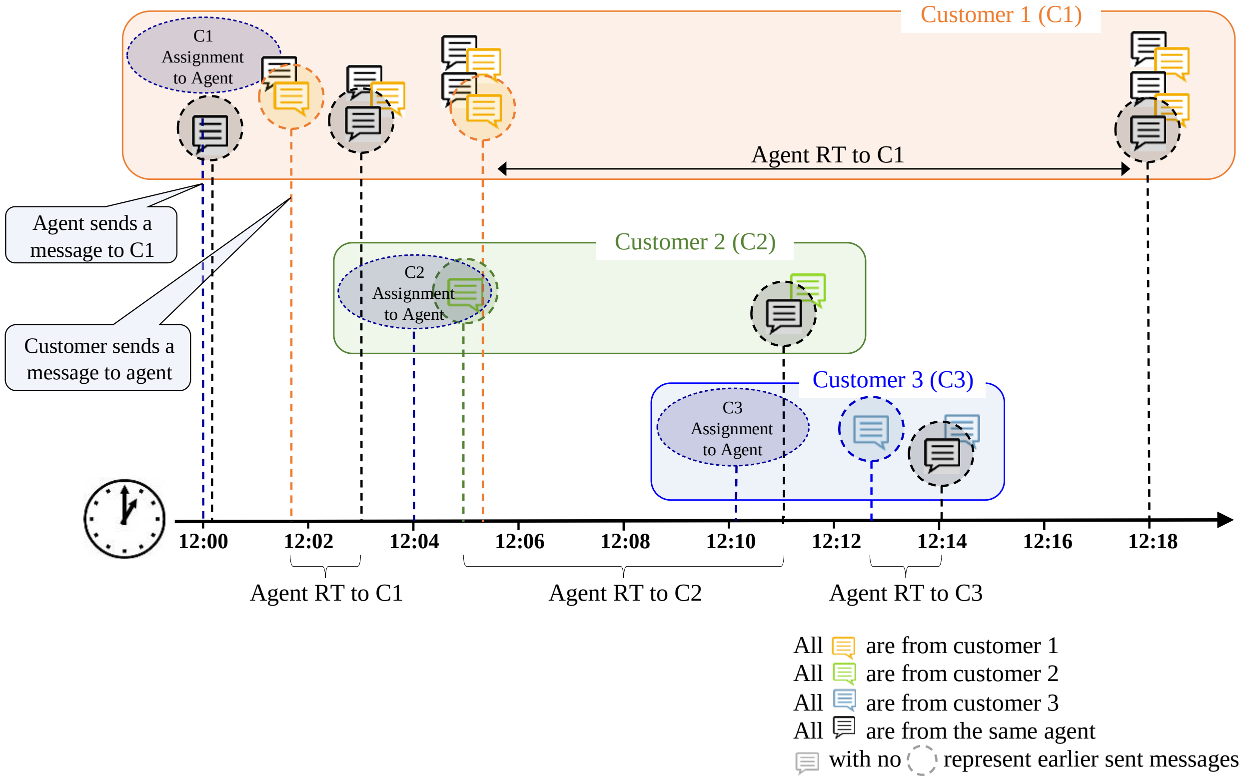
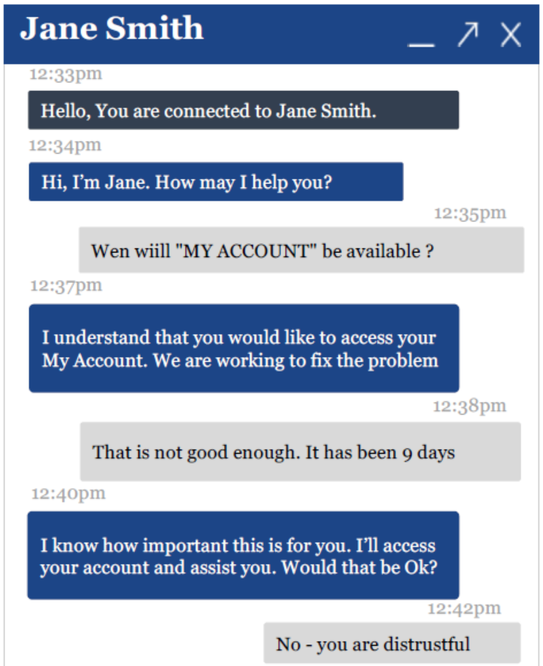


Figure 3 - Schematic View of Simultaneous Agent Chats with Three Customers.

Each conversation is identified by a chat ID, agent ID, date, issue (sales or service), and time the customer waited in a queue before chatting. For each line within a chat our data contains the following: a time-stamp of when the line was sent, notation of who wrote the line (customer or agent), number of words, and an emotion score. To ensure privacy, our data do not include the text of the conversations or any demographic indicators of customers or agents.

4.1.3 **Measuring Operational Features of Conversations**

Figure 4 provides an example of a chat, its recorded data, customer emotion, and two computed variables: agent RT and number of turns. *Agent RT* is computed as the elapsed time between each customer message and the agent response. The number of turns is computed as the total number of customer-agent iterations.



Author	Agent Response Time (RT)	Number of Words	Emotion	Turn
System	-	-	-	-
Agent	-	8	-	1
Customer	-	6	0 (neutral)	2
Agent	75 sec	18	-	3
Customer	-	10	-1 (negative)	4
Agent	82 sec	19	-	5
Customer	-	4	-1 (negative)	6
Total Number of Turns	-	-	-	6

Figure 4 - An illustration of Agent-Customer Chat and Measures.

We compute agent RT (rather than service time or customer length of stay (LOS)) because (1) agent RT translates directly into agent efficiency; (2) agent RT defines the customer wait time experience, which service delivery must minimize; (3) agent RT is free from endogenous time intervals, such as customer RT, so is preferable to LOS. We note that due to concurrency, agent RT is a result of tasks being performed for a focal customer and for other customers. Customers are generally blind to agent's work processes. Customers can see a note indicating when the agent is typing to them; but we do not have the records of such notes, so could not include this in our analyses. Hence, our data does not allow a "clean" decomposition of agent RT into focal customer service time and service times to other customers. Therefore as a proxy for agent effort, we use the number of words in each agent message, similar to Goes et al. (2018). Using the meta-data described, we calculate the number of concurrent customers assigned to each agent, and control for agent multitasking by including this concurrency measure. We also control for the operational workload, using the number of customers waiting in the queue. This information is also available on the agent screen.

4.1.4 Measuring Customer Emotions in Conversations

We measure customer expressions of positive and negative emotions as two sides of a single scale (Fredrickson & Kahneman, 1993; Gabriel & Diefendorff, 2015), and use the terms "emotion" and "sentiment" interchangeably to refer to customer expressions of emotion. We reviewed multiple Sentiment Analysis tools (e.g., Linguistic Inquiry Word Count (LIWC*)*, Tausczik and Pennebaker (2010), *SentiStrength*, Thelwall (2013), *CustSent*, Yom-Tov et al. (2018), and *Sentiment Tree bank*, Socher et al. (2013)), and selected two tools–*SentiStrength* and *CustSent*–that offer the most accurate assessments of customer emotion in chat service. *SentiStrength* was developed to assess positive and negative emotion in short texts, and *CustSent* was designed to analyze sentiment in customer service conversations. Both tools utilize labeled dictionaries coupled with Natural Language Processing techniques, and have better accuracy than other tools in the customer service context: Yom-Tov et al. (2018) reports that *SentiStrength* has the highest recall, and *CustSent* has the highest precision values with customer service texts (see Appendix 2 for recall and precision data of the tools).[3]

These two tools assign a valence and intensity value for the emotion expressed in a message. Negative and positive signs represent negative and positive emotions, respectively. The

[3] Following reviewer queries, we considered also measuring agent emotions. We used the same tools to analyze agent sentiment, searching for instances of agent expression of negative emotion in a sample of about 200 agent messages. We found that all agent messages that express what the tools construed as negative emotions include some version of apology (e.g.,"I am so sorry you had to wait.") or reassurance (e.g., "Don't worry, we'll find the annoying mistake"). We found no agent messages with negative emotions such as anger or frustration. Thus, the sentiment score of agent messages is qualitatively different from the sentiment score of customer messages. We therefore do not develop hypotheses about agent emotion. We included agent sentiment in our analyses of the robustness tests, and the results did not change much (see sensitivity analysis in Section 4.4.3).

score itself indicates the intensity of the emotion. *SentiStrength* sentiment scores range from -4 to +4. For example, the following text received a score of +1:

"That enabled me to access my account. Thanks, that's really helpful."

In contrast, the following text received a score of -1:

"I don't know. I'm concerned about my credited miles."

CustSent has no hard limits on the sentiment scores, but in our data these scores range from -12 to +10. As reported below, to reduce measurement error we combine the scores of the two tools in our analyses (see Section 4.3).

Figure 5 describes the customer emotions evaluated by *SentiStrength* in our data. Figure 5(a) shows the proportion of chats having only positive emotion, only negative emotion, multiple emotion (both positive and negative) and neutral. More than 85% of chats include emotion, which positions emotion as a central feature of service. Figure 5(b) shows the proportion of customer messages that contain positive, negative or neutral expressions, and suggests that most messages within conversations are neutral. Both the chat and message analysis show that positive emotion is more commonly expressed than negative emotion. Figure 5(c), which shows the distribution of emotion intensity in messages, further confirms the higher prevalence of positive emotion.

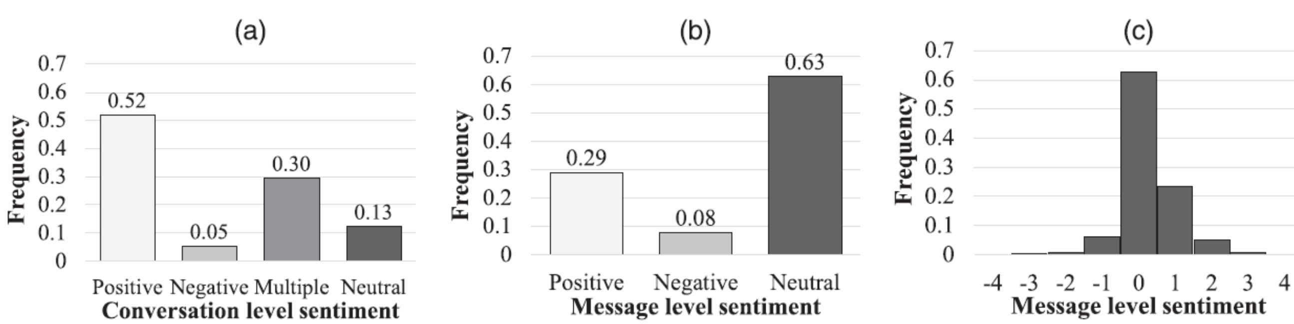


Figure 5 - Frequency of Emotion in Customer Service (SentiStrength Score).
Notes. (a) Full service conversations. (b) Customer messages. (c) Sentiment distribution

Figure 6 graphically depicts the association between customer emotion and agent RT, showing a kernel smoothing of average agent RT (throughout the chat) as a function of average customer emotion. The apparent relationship between emotional workload and agent RT is analyzed below by testing the causal effects in this relationship, while controlling multiple relevant factors. Next we formulate hypotheses that relate agent behavior and customer emotion and then test these hypotheses with an econometric framework in Section 4.3.

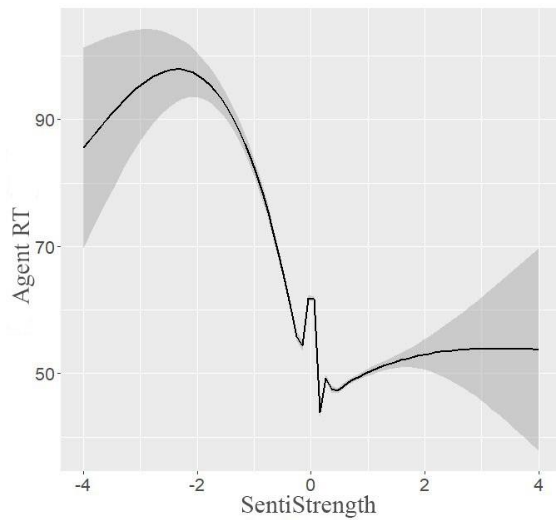


Figure 6 - Covariation of Agent RT and Customer Sentiment. Marked in gray is the 95% Confidence Interval.

4.2. **Theory Development**

Operational and marketing perspectives typically consider customer emotions as responses to agent behavior and as indicators of customer satisfaction; mere outcomes of an interaction. In contrast, we view customer emotion as a unique source of load for service agents, and propose that such emotional workload influences agent performance-related behavior, and specifically agent RT. Below, we first review literature that supports our emotional workload theory and then we discuss the opposite, and more prevalent, view that agent RT impacts customer emotion.

4.2.1 **Effects of Customer Emotions on Agent Behavior**

The Episodic Model of Affect and Performance (Weiss & Cropanzano, 1996), positions work as a series of episodes in which emotional experiences vary, and influence work performance (Beal, Weiss, Barros, & MacDermid, 2005). The model suggests that emotional events at work (e.g., exposure to an angry customer), influence service agents, because they affect mental resources. This model is the foundation for our predictions. Experimental research has shown that customer rudeness and anger hamper service agents' performance of various tasks (Rafaeli et al., 2012), due to disruption of cognitive processes (Porath & Erez, 2007). To illustrate, participants in a simulation of customer service work erred more when processing customer requests phrased in a hostile manner than when requests were phrased politely (Goldberg & Grandey, 2007). Similarly, a series of lab studies showed that listening to verbally abusive customers hampered participants' ability to recall the content of the conversation (Rafaeli et al., 2012). Building on this, we expect that agents need extra time to resolve customer issues expressed with negative emotion.

Moreover, agents who encounter customer emotions must often suppress their own genuine emotions, and display organizationally appropriate responses (Geddes & Callister, 2007), performing the demanding task known as "*Emotional Labor*" (Grandey et al., 2010; Rafaeli & Sutton, 1987). The additional effort required to convey appropriate emotions likely requires extra time from the agent (Sutton & Rafaeli, 1988). In this vein, when customer emotion is *positive*, agent emotion corresponds to the appropriate response and so no extra effort is required for the agent to express their response. Additionally, positive customer emotion is replenishing, and improves agent motivation and available cognitive resources (Bakker & Demerouti, 2007), both of which help agents solve customer issues more rapidly. Hence, our first hypothesis:

Hypothesis 1. *The more negative emotion a focal customer expresses in a given message, the longer the agent RT in the subsequent agent message.*

Although multiple mechanisms could explain Hypothesis 1, we test what we see as a key mechanism–agent effort. Customer messages that include negative emotion, require additional communication effort (compared to positive/neutral messages), in addition to the effort required to generally resolve the customer issue (Geddes & Callister, 2007). For example, agents must acknowledge customer frustration or dissatisfaction and may need to apologize to customers. These additional communication efforts will lengthen the agent text. Hence, our second hypothesis:

Hypothesis 2. *The more negative emotion a focal customer expresses in a given message, the larger the agent effort in the subsequent agent message.*

We position agent effort as a mechanism through which customer emotion influences agent RT, suggesting that agent effort acts as a mediator. Hence our next hypothesis:

Hypothesis 3. *Agent effort mediates the effect of customer emotion on agent RT.*

The hypotheses presented so far are relevant to the message-level (i.e., messages within conversations). A conversation-level analysis is relevant for considering the effects of customer emotion on the length of a conversation. Two effects confirmed by Rafaeli et al. (2012), Porath and Erez (2007) and others suggest that negative customer emotions will prolong a service conversation. Customer expressions of negative emotions hamper agents' cognitive processing and increase agent errors, which extend the length of a conversation. In addition, customers' negative emotions distract agents, leading to more agent inquiries as the agent seeks to understand the customer needs. Hence, our next hypothesis:

Hypothesis 4. *The more negative emotion a focal customer expresses during a conversation, the greater the number of turns required to complete the conversation.*

We note that a competing hypothesis for Hypothesis 4, as suggested by Sutton and Rafaeli (Sutton & Rafaeli, 1988), could be that customer expressions of positive emotion create more customer engagement, and extend the service conversation. Similarly, positive emotions create a more pleasant work environment, and may motivate agents to spend more time (and thus exchange more messages) in the conversation. But this competing theory does not have strong support in empirical research.

4.2.2 **Effects of Agent Behavior on Customer Emotions**

Another side of the customer-agent interaction is the influence of agent RT on customer emotions. Customers can construe an agent's RT as wait time, leading to expressions of negative customer emotion. Customers dislike waiting (Larson, 1987; Maister, 1985; Taylor, 1994), so much so that people waiting often abandon a service (Allon, Federgruen, & Pierson, 2011; Mandelbaum & Zeltyn, 2013). Importantly, agent RT in chat service can include delays and in-service waits due to concurrency of other customers (Goes et al., 2018). This can create unexplained waiting which may annoy and frustrate customers, evoking expressions of negative emotion. Hence, our next hypothesis:

Hypothesis 5. *Longer agent RT to a focal customer message creates more negative emotion in the subsequent customer message.*

Another way that agents might influence customer emotion is by investing effort. Customers are more satisfied when they feel that an agent works harder to resolve their issue (Groth, Hennig-Thurau, & Walsh, 2009). Customers seek specific cues to assess the agent effort, and cues such as the time and energy an agent spends on a customer can impact perceived effort above and beyond the outcome of the service (Mohr & Bitner, 1995). In the context of chat-service, customers are detached from the service agent and cannot see when an agent is working toward solving their inquiry. However, customers can perceive invested agent time and effort through the length of an agent's message. If the number of words an agent writes is indeed a proxy for the customer's perception of agent effort, then we would expect that customers who encounter long messages will be more satisfied, and hence will express more positive emotions. In the same spirit, shorter messages would signal reduced agent effort and lead to customer dissatisfaction and the expression of more negative customer emotions. Hence, our next hypothesis:

Hypothesis 6. *An increase in agent effort (the number of words in an agent message) creates more positive customer emotion in the subsequent message.*

Finally, effects on customers may also accrue as a conversation unfolds. Customers participate in a service conversation to accomplish a goal or a set of goals. Long service conversations can make customers frustrated (Katz, Larson, & Larson, 1991) and angry (Casado Díaz & Más Ruíz, 2002). When a service conversation is very long, customers may strategically express “fake” anger, to signal dominance and toughness (Knutson, 1996; Tiedens, 2001). Customers can also perceive long service times as unprofessional (Anand, Paç, & Veeraraghavan, 2011; Casado Díaz & Más Ruíz, 2002), since a longer conversation might signal that the agent is unable to solve the customer’s problem. A sense of unprofessional service can translate into customer expressions of negative emotion. Hence, our final hypothesis:

Hypothesis 7. *As the number of turns within a conversation increases, customers express more negative emotion.*

Next, we empirically examine our hypotheses and decipher the complex relationship between customer expressed emotion and agent behavior.

4.3. Econometric Specification

This section develops an econometric framework to test the causal effects our hypotheses predict. An important challenge in the estimation arises from omitted factors related to the complexity of a focal case handled by an agent. Cases of higher complexity are likely to be associated with longer agent RTs, because they require more effort to handle. More complex cases are also likely to evoke negative customer emotion. We can include some observable proxies of case complexity in the model, but there are dimensions of complexity which cannot be measured and therefore become confounds that can bias the estimates. A second complication is reversed causality between agent behavior and customer emotion, whereby longer agent RTs may enhance customer frustration. This produces a simultaneity problem between customer emotion and agent behavior: cases that take longer to handle also tend to have negative customer emotions, and the causal relationships between the two is not clear (Manski, 1993).

The empirical strategy we used to identify the causal effect of emotion on agent behavior is to exploit the panel structure of the data, using variation across the sequence of messages within a conversation as a source of identification. Let *i* index the customer-agent conversation associated to a case and let $NTurns_i$ denote the number of turns, with t=1...$NTurns_i$ representing each turn within that conversation. The variable EMO_{it} measures the emotion of a customer message in turn t, and RT_{it} the agent response time to a message t. RT is modeled as:

$$\log(\mathrm{RT}_{\mathrm{it}})=\delta_{\mathrm{i}} +\beta EMO_{\mathrm{it}-1} +\gamma W_{\mathrm{it}} +\tau ConvStage_{\mathrm{it}} +u_{\mathrm{it}}, \quad (1)$$

where δ_i is a fixed effect for the conversation, W_{it} are workload related factors that vary during the conversation and u_{it} is an error term. The coefficient of interest is β, which we predict to be negative according to Hypothesis 1. Other applications with similar data revealed that *EMO* has a positive trend during a conversation (Yom-Tov et al., 2018). To account for this trend, the covariate $ConvStage_{it}=t/NTurns_i$ is specified to capture the stage of conversation i where the focal turn t occurs. This control variable is included in the econometric models that are analyzed at the message (*it*) level.

The fixed effect δ_i controls for several unobserved factors that could lead to omitted variable bias. In particular, it captures the complexity of a case, which by definition does not vary during the case conversation. Because conversations last on the order of minutes (11.7 minutes on average), effects due to day of the week and hour of the day are also captured by δ_i. Because a conversation is handled by a single agent, all agent related factors are also absorbed by the fixed effect.

Previous work showed several mechanisms that relate workload to agent productivity (for a review see Delasay et al. (2019)). Workload can affect the speed of an agent's work by leading to fatigue, thereby reducing productivity and compliance with process standards (Dai, Milkman, Hofmann, & Staats, 2015). On the other hand, current and pending workload can put pressure on an agent to work harder and increase productivity (Kc & Terwiesch, 2009; Tan & Netessine, 2014). In settings with a shared queue among multiple agents, social loafing can lead agents to slow down when facing a long queue (Wang & Zhou, 2016). To capture the effects of the customer queue, a covariate measuring the number of customers in the queue at the beginning of the RT interval, $NumInQueue_{it}$ is included as a control.

Agents in chat contact centers can simultaneously handle multiple conversations, a workload that can also create fatigue and pressure effects. Handling simultaneous conversations is a form of multitasking, known to also affect productivity (Bray, Coviello, Ichino, & Persico, 2016; Goes et al., 2018; Kc, 2014). The number of concurrent chats ($Concurrent_{it}$) is measured as an average during the RT interval. Given the dynamics of work assignment in contact centers, both *NumInQueue* and *Concurrent* can vary substantially during the course of a conversation, but are not controlled by the agent and are therefore considered exogenous; these two variables are the main covariates included in W_{it} (we also consider alternative measures of concurrency in the Sensitivity Analysis in Section 4.4.3). Other workload related effects, such as the hours elapsed during the working shift, do not vary much during a conversation due to its relatively short duration, and are therefore absorbed in the fixed effect δ_i.

Identification in this model is driven by the variation in emotion across customer messages during the same conversation. Recall that one of the concerns regarding the identification of the causal effect of emotion on agent behavior was reverse causality: it is possible that *EMO* and *RT* affect each other. Our regression Model (1) exploits the sequencing of the messages to avoid this reverse causality. The variable RT_{it} is measured *after* the customer expresses emotion in his/her message in turn $t-1$, hence it could not have influenced EMO_{it-1}. Furthermore, the detailed model–including the conversation fixed-effect δ_i–controls for most of the omitted variables related to case and agent heterogeneity, providing a clean identification strategy.

A final concern for identification is measurement error in the *EMO* variable, which could lead to an attenuation bias in the associated coefficient. We started with one measure of *EMO* for the analysis *SentiStrength* (Thelwall, 2013), and used a second measure, *CustSent* (Yom-Tov et al., 2018) to mitigate concerns about measurement error in the first measure. There are differences between the two measures, both in the dictionaries they use and in the range of sentiment scores. Nonetheless, the two measures are highly positively correlated (r=0.63, p<0.001). Hence we use the second *CustSent* emotion measure as an Instrumental Variable (IV), which eliminates measurement error (Wansbeek & Meijer, 2003), in all of the models with *EMO* as an independent variable.

4.3.1 Decomposing the Effect of Customer Emotions on Agent Behavior

The effect of customer emotion on agent RT can be direct or indirect (mediated) through agent effort (see Hypotheses 1 and 3). We use the number of words in agents' responses as a proxy for agent effort, similar to (Goes et al., 2018). The number of words in an agent message ($NumWords_{it}$) is included in the specifications as follows:

$$\log(RT_{it})=\delta_i+\beta_1 EMO_{it-1}+\beta_2\log(NumWords_{it})+\gamma W_{it}+\tau ConvStage_{it}+u_{it}; \qquad (2)$$

$$\log(NumWords_{it})=\delta_i+\beta_3 EMO_{it-1}+\gamma W_{it}+\tau ConvStage_{it}+u_{it}. \qquad (3)$$

This specification, which includes the same control variables as Model (1), captures the direct and indirect effects of *EMO* on *RT* (Hypotheses 1 and 3). Coefficient β_3 captures the effect of customer emotion on agent effort in responding to the customer (Hypothesis 2), using *NumWords* as a proxy for effort. This effect translates into an impact on *RT* because longer text requires more time to write ($\beta_2 > 0$). The coefficient β_1 captures other effects of emotion on *RT*, since *NumWords* may not be a perfect proxy. Thus, β_1 may include other effort-related aspects not reflected in a longer agent message (e.g., scrutinized search in the Customer Relationship Management software). We do not have documentation of agents' activity outside of the chat platform, and therefore cannot measure the full agent effort directly. Despite these limitations, Models (2) and (3) provide information about the alternative paths through which customer emotion affects agent behavior. As before, we correct for measurement errors and include the other control variables (with some abuse of notation, the same parameters are used for the controls' coefficients to facilitate reading).

Models (2) and (3) correspond to a mediation model where the effect of *EMO* on the agent *RT* can be decomposed into a direct effect (coefficient β_1) and an indirect effect through *NumWords* (measured by $\beta_3 \times \beta_2$). A key assumption to identify the coefficients $\beta=(\beta_1, \beta_2, \beta_3)$ is that u_{it} and v_{it} are independent, that is, unobservable factors that affect *NumWords* do not directly affect *RT* (conditional in all the controls of the model). Recall that the models include conversation fixed effects, which control for the case complexity and customer and agent characteristics; these controls are needed to justify this identification assumption. Under these conditions, Models (2) and (3) can be estimated as independent regression models (using IVs to mitigate the measurement error of *EMO*) to provide consistent estimates of the model parameters. However, calculating confidence intervals for the indirect effect $\beta_3 \times \beta_2$ is complicated because the two estimators are correlated due to sampling error. We use a bootstrapping approach to estimate the models and compute confidence intervals, using the methods developed by Hayes and Rockwood (2020) to conduct mediation analysis with panel data.

4.3.2 Effect of Customer Emotions on the Length of a Conversation

Estimating the effect of emotions on the number of turns in a conversation requires a different approach. We propose here two identification strategies (Models (4) and (5)). First, we model the number of turns as a static variable that we measure at the conversation level. In this case, the basic unit of analysis is a conversation. The model includes the effect of emotion in the first customer message of the conversation, EMO_{i1}. According to Hypothesis 4, we expect that the coefficient of EMO, β_4, to be negative. One may be tempted to average the emotion across all messages in a conversation, but this is problematic due to the reverse causality problem discussed earlier: customer emotion affects agent behavior but agent behavior also affects customer emotion. Furthermore, measuring the impact of the customer emotion in the first message can be useful for balancing work allocation between agents (see Section 4.5). To account for agent workload we include average concurrency during a conversation (*Concurrent*), and the number of customers in queue when the conversation started (*NumInQueue*). Both indicators are exogenous, therefore, reversed causality is not a concern here. We use the following regression model to estimate the impact of customer emotion on the number of turns:

$$NTurns_{i=}\rho_{a(i)} + \beta_4 EMO_{i1} + \gamma W_i + \psi X_i + w_i. \qquad (4)$$

The term $\rho_{a(i)}$ is a fixed effect of the agent serving chat i and w_i is an error term. The other covariates in Model (4) are discussed next. Since the model is estimated with a cross-section of conversations, it is important to control for case complexity. The number of words in the first customer message ($CustWords_1$) is an exogenous variable used to proxy the complexity of the case, included as a covariate with log transformation (to keep consistency with the previous models). To capture seasonal effects, a weekday-weekend dummy and hour of the day dummies are included (*IsWeekend* and *HourOfDay*, respectively). The type of service case (*SrvType*) is controlled through a dummy variable. Finally, changes in agent behavior due to fatigue are controlled with dummy variables for each hour worked during the shift (*ShiftTime*). These covariates are included in the set of controls denoted by X_i.

As predicted by Hypothesis 4, customer emotion during the conversation can also affect the extension of the conversation, as measured by the number of turns. For this we use an alternative identification strategy. Consider a discrete-time duration model, where “periods” are represented by each turn in a conversation. Define y_{it} as a binary dependent variable which is equal to one if conversation i ends in turn t, and zero otherwise. The $\Pr(y_{it=}1)$ can be viewed as a hazard rate of the length of a conversation and can be modeled through a Probit model as:

$$\Phi^{-1}(\Pr(y_{it}=1))=\rho a(i) + \beta_5 EMO_{it-1} + \beta_6 EMO_{i1} + \gamma W_{it} + \varphi Xi, \qquad (5)$$

where W_{it} are the workload-related variables included in Models (1), (2) and (3), and $\Phi^{-1}(\cdot)$ is the inverse of the standard normal distribution. Since there is only one spell of messages for each conversation, this model cannot include conversation fixed effects (because $y_{it=}1$ only for the last turn of each conversation i). Therefore, the same control variables X_i from Model (4) are included in this model, to capture cross-sectional differences across conversations. Additional specifications were estimated including the emotion in the first customer message, EMO_{i1}, as a proxy for potential observable factors that could be correlated with the initial emotion of each conversation. The coefficient of interest in Model (5) is β_5, which measures the impact of customer emotion in the previous message on the hazard rate (likelihood of terminating the conversation); Hypothesis 4 predicts a positive effect, $\beta_5 > 0$–the more positive the emotion, the shorter the conversation should be and the probability that the conversation will end in the next turn should increase. As before, *EMO* was instrumented in Models (4) and (5) to reduce measurement error.

4.3.3 Modeling the Effect of Agent Behavior on Customer Emotions

Our next hypothesis regards the influence of agent behavior on customer emotion (Hypothesis 5). The empirical strategy we used to test this hypothesis is as follows. First, we consider the following specification to estimate the effect of agent RT on customer emotion:

$$EMO_{it}=\delta_{i} + \alpha \log(RT_{it-1}) + \tau ConvStage_{it} + e_{it}. \qquad (6)$$

The unobservable e_{it} includes the quality of the agent response as perceived by the customer, which is difficult to control with the variables observed in the data. It is then plausible that agent RT is positively correlated with the quality of the agent response, since agents need to do time consuming work to properly address a customer issue. This positive correlation between *RT* and the error term induces a positive bias in the estimation of α. Our approach to correct for this bias is to use IVs that affect agent RT but do not *directly* affect customer emotion. Recall from Model (1) that *RT* is affected by the agent workload, W_{it}. In the context of this application, customers cannot directly observe the workload of the agent, thereby the effect of this workload can only affect emotion through the RT perceived by the customer. Measuring the effect of *RT* induced by variation in an agent's workload is also useful from a managerial perspective, as it provides insights on how workload management and staffing decisions can affect customer emotion. According to Hypothesis 5, we expect the coefficient α to be negative. Model (6) can be further refined by including additional factors associated with agent effort,

specifically, *NumWords* and *Turn*:

$$EMO_{it=}\delta_{i} + \alpha_{1}\log(RT_{it-1})+ \alpha_{2}\log(NumWords_{it-1})+ \alpha_{3}Turn_{it} + \tau ConvStage_{it} + e_{it}. \qquad (7)$$

The number of words (*NumWords*) in a message (our proxy for agent effort), is directly observable by the customer. Longer agent messages might be perceived by customers as increased agent effort, thereby generating positive emotion (see Hypothesis 6). We therefore expect the coefficient α_2 to be positive. As noted, customers cannot see all the activities performed by an agent during the *RT*, and may therefore interpret a long *RT* as lack of agent dedication, which would produce negative customer emotion. According to Hypothesis 7, customer emotion can also be affected by an extension of the conversation, which is captured through the variable $Turn_{it}$ (i.e., the ordinal count of turns in a conversation). Notice that $Turn_{it}$ and $ConvStage_{it}$ are correlated but not perfectly co-linear, hence their effects can be identified separately and with reasonable precision given the large sample size. A potential issue is that *RT*, *NumWords* and *Turn* can all be correlated with the complexity of the customer issue, since more complex issues require more effort from the agent and a longer conversation. But recall that the fixed-effect δ_i controls for case complexity, mitigating this omitted variable bias. As before, *RT* is instrumented with the workload-related exogenous variables *W* (*Concurrent* and *NumInQueue*) in order to mitigate the endogeneity bias that can be generated by unobservable quality of the agent's response.

Table 2 summarizes the variables used in all the econometric models. The next section discusses further specification details, summary statistics and the estimation results.

4.4. **Estimation Results**

Table 3 reports summary statistics of the variables used in the estimation. The top panel shows the variables included in Models (1)–(3) and (5)–(7), with messages as the unit of analysis and the bottom panel shows variables of Model (4), with conversation as the unit of analysis. In both cases outliers were removed from the sample, in order to avoid influence of extreme cases on the estimation. The Max column indicates the cutoffs used for excluding outliers. In the message-level data, we defined outliers as observations with *RT* below the 5th percentile (below 8 seconds) and above the 95^{th} percentile (above 1641 seconds). We removed observations where *NumWords* was above the 95^{th} percentile (387 words). We also removed conversations with data errors in the *ShiftTime* and conversations that were conducted after the eighth hour of an agent's shift, to focus only on regular shifts (95% of conversations). The elimination of outliers and chats with missing data removed a total of 75,160 conversations from the analysis, leaving an effective sample size of 141,654 chats. Appendix 3 and 4 show the inter-correlation among the variables. As a robustness check, all analyses were replicated with the outliers included in the sample (see Sensitivity Analysis in Section 4.4.3).

Table 2 - Labels and Coding of Study Variables

Variable	Description and coding
Dependent variables	
RT_{it}	Agent response time to a focal customer message in turn t of conversation i [seconds]
$NumWords_{it}$	Number of words agent wrote to a focal customer in turn t of conversation i (a proxy of agent effort)
$NTurns_i$	Number of iterations between customer and agent in conversation i (an iteration is counted when one party answers the second party)
EMO_{it}	Customer emotion in turn t of conversation i as measured by *SentiStrength*
W variables:	Agent workload
$NumInQueue_{it}$	Number of customers in queue at the beginning of turn t of conversation i
$Concurrent_{it}$	Weighted average of number of parallel chats handled by agent during turn t of conversation i
X variables:	Complexity of problem and time variables
$SrvType_i$	Type of service in conversation i: support (coded 0; 50.81%) or sales (coded 1)
$CustWords_{it}$	Number of words customer wrote in turn t of a conversation i
$ShiftTime_i$	Time that passed since an agent started the shift until the beginning of a conversation i [hours]
$HourOfDay_i$	Hour (8:00-23:00) of the conversation i
$IsWeekend_i$	Weekday: Mon-Fri (coded 0; 72.24%), Weekend: Sat-Sun (coded 1)
Other variables:	
$Turn_{it}$	Ordinal number of current turn t in a conversation i
$ConvStage_{it}$	Progress of conversation completed (Range 0-1)
$CustSent_{it}$	Second measure of customer emotion in turn t of conversation i

4.4.1 **Effect of Customer Emotions on Agent Behavior**

Table 4 shows the estimation results of econometric Models (1), (2), and (3). Recall that Models (1)–(3) are at the message-level of analysis, and include fixed effects of the conversation, so the coefficients are estimated using variation across turns of each conversation. Models (2) and (3) are estimated using a mediation model based on Hayes and Rockwood (2020), using bootstrapping to compute the standard errors[4].

[4] This method is designed especially for panel data, drawing conversations with replacements from the data in the resampling procedure. For each re-sample, Equations (2) and (3) are estimated separately using 2SLS, which accounts for the nested nature of the data. Confidence intervals are calculated based on the empirical distribution of the estimates from each re-sample. See Hayes (2018) for a description of this bootstrap process.

The results of Model (1) confirm a negative and statistically significant effect of *EMO* on *RT*, supporting Hypothesis 1. The key covariate for Model (1) is *EMO*, instrumented with *CustSent*. The effect is substantial, with a one point improvement in customer emotion (i.e., emotion becomes more positive) reducing *RT* by 20.6% (a 14 second reduction in average RT per message).

Other control variables also have significant effects on *RT*: *Concurrent* has a positive effect, meaning that simultaneous conversations with multiple customers increase the RT to each focal

Table 3 - Descriptive Statistics of Study Variables

Variable	Mean	Median	SD	Min	Max
Message level (N=650,856)					
EMO [$SentiStrength$]	0.27	0.00	0.74	-4	4
$CustSent$	0.22	0.00	0.72	-12	10
RT [seconds]	65.25	47.00	66.10	8	1641
$log(RT)$	3.84	3.85	0.80	2.08	7.4
$NumWords$	34.58	27.00	26.16	1	387
$log(NumWords)$	3.30	3.30	0.72	0	5.96
$Concurrent$	2.33	2.47	0.72	1	3
$NumInQueue$	2.52	1.00	3.87	0	73
$ConvStage$ [%]	0.58	0.58	0.27	0.02	1
$ShiftTime$ [hours]	3.63	3.41	2.31	0	8.16
$Turn$	8.78	6.00	7.99	2	132
Conversation level (N=141,654)					
$NTurns$	10.18	8.00	7.01	2	114
$log(NTurns)$	2.14	2.08	0.6	0.69	4.74
EMO_1 [$SentiStrength$]	0.10	0.00	0.62	-4	4
$CustSent_1$	-0.04	0.00	0.51	-10	7.5
$Concurrent$	2.44	2.65	0.58	1	3
$NumInQueue$	3.13	2.00	4.00	0	72
$CustWords_1$	26.82	23.00	19.49	1	1131
$log(CustWords_1)$	3.00	3.14	0.88	0	7.03
$ShiftTime$ [hours]	3.47	3.26	2.27	0	7.76
$HourOfDay$	14.23	14.00	3.75	8	22

customer. *NumInQueue* has a small positive effect, meaning that a longer queue of customers makes agents work slightly slower. The effect of *ConvStage* is positive and suggests an increase in *RT* toward the end of conversations.

The second and third columns of Table 4 show the results for estimates of the mediation Model (2)–(3). Supporting Hypothesis 2, Model (3) shows that *EMO* increases log(*NumWords*), though the magnitude of the effect is small. Model (3) also shows a negative effect of concurrent conversations: as agents increase multitasking, they write shorter messages to each customer.

Model (2), with log(*RT*) as a dependent variable, includes *EMO* and the logarithm of *NumWords* as the main variables of interest. *EMO* has a negative and significant effect on *RT*, similar in magnitude to the estimates of Model (1): a one point increase in *EMO* reduces *RT* by 19.7%. The number of words in the message *NumWords* has a large *positive* effect on *RT*, which is expected because a longer text takes more time to write. Doubling the length of an agent's message increases *RT* by 44.6%. The bottom panel shows the indirect effect of *EMO* on log(*RT*), with the significant mediation of log(*NumWords*), supporting Hypothesis 3. Overall, a one point increase in *EMO* causes a 19.7% direct reduction in *RT* plus an indirect effect (through *NumWord*) that increases *RT* by 0.30%.

The effect of the other covariates in Model (2) are similar to those reported for Model (1), except for *ConvStage* which now has a smaller magnitude: from 0.246 to 0.006. The longer RTs toward the end of the conversation appear to be partially explained by the length of the messages: Model (3) suggests that agent messages tend to be longer as the conversation progresses.

Table 4 - Effect of Customer Emotion on Agent Behavior (Outliers Excluded, EMO_{t-1} is Instrumented using $CustSent_{t-1}$)

	Model(1) $log(RT)$	Model (3) $log(NumWords)$	Model (2) $log(RT)$	Model (3) $log(NumWords)$	Model (2) $log(RT)$
EMO_{t-1}	-0.206*** (0.0028)	0.007** (0.0024)	-0.197*** (0.0024)		
$EMO\_positive_{t-1}$				0.020*** (0.002)	-0.153*** (0.002)
$EMO\_negative_{t-1}$				0.043*** (0.004)	0.042*** (0.004)
$Concurrent_t$	0.057*** (0.0026)	-0.040*** (0.0024)	0.074*** (0.0024)	-0.040*** (0.002)	0.074*** (0.002)
$NumInQueue_t$	0.003*** (0.0007)	0.002* (0.0006)	0.003*** (0.0006)		
$ConvStage_t$	0.246*** (0.0043)	0.464*** (0.0037)	0.006† (0.0004)	0.458*** (0.004)	-0.033*** (0.004)
$log(NumWords_t)$			0.446*** (0.0014)		0.444*** (0.001)
$NumInQueue$ (chat level)				0.001*** (0.001)	0.001* (0.001)
Conversation Fixed Effect	Included	Included	Included	Included	Included
Constant	3.616*** (0.0070)	2.764*** (0.0064)	1.809*** (0.0099)	2.753*** (0.006)	1.831*** (0.01)
Indirect Effects					
EMO via $log(NumWords)$			0.003*** (0.0011)		
$EMO\_positive$					0.009*** (0.001)
$EMO\_negative$					0.019*** (0.002)
Observations	650,856	650,856	650,856	650,159	650,159

Standard errors in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ †$p < 0.1$
Note. The Macro we use in the mediation analysis (Hayes and Rockwood 2020) allows up to 3 message-level covariates and up to 3 chat-level covariates. This means that for the estimation with categorical emotion (last two columns) we had to aggregate $NumInQueue$ to the chat-level. This produced some additional missing values, and therefore reduced N size by 697 observations.

The last two columns of Table 4 replicate the estimation of Models (2) and (3) but include three categories of customer emotion: Negative (*EMO* < 0), Neutral (*EMO* equals to zero) and Positive (*EMO* >0). The Neutral category is the excluded dummy. Model (3) suggests that positive customer emotions have a small positive effect on *NumWords*: messages with positive emotion (compared to neutral emotion) increase the number of words written by the agent by 2%. Similarly, negative customer emotion (compared to neutral emotion) increases *NumWords* by 4.3%. Overall, the impact of customer emotion on the length of agents' messages is relatively small, but there is evidence that agents put in more effort when customers express emotions and the effort is greater when the emotion is negative.

The fifth column in Table 4 shows the estimation of Model (2) with categories of customer emotion, and confirms that the effect of customer emotion is negative, monotone, and economically significant. Messages with negative emotion receive RTs which are about 20% longer relative to messages with positive emotion. These results are consistent with the linear specifications. Interestingly, the largest effect is observed in *EMO positive* when positive customer emotion reduces *RT* by 15.3% compared to neutral emotion.

We next discuss the estimation of Models (4) and (5), which assess the effect of *EMO* on the length of conversations as measured by the number of turns. The first column of Table 5 shows the estimation of Model (4), using a cross section of conversations and EMO_1–the emotion of the first customer message in the conversation–as the main covariate of interest (which is instrumented to reduce attenuation bias due to measurement error). The coefficient of EMO_1 is negative and statistically significant, where a one point reduction in customer emotion increases the number of turns in the conversation by 1.684, equivalent to 17% of the mean, which is economically significant. This result is aligned with those obtained in Models (1)–(3), providing further support that conversations with negative customer emotion tend to be longer and require more time from the agent. In terms of the other covariates, a higher number of concurrent customers handled by the agent during the conversation reduces the number of turns required to finish the case, suggesting that agents may be speeding-up to close cases faster when their workload is high. The number of words in the first customer message has a negative effect on the number of turns, and the effect of the number of customers in queue is small.

The second and third column in Table 5 report the estimates of two specifications of the hazard Model (5): they both include EMO_{t-1}–customer emotion in the previous message–but the third column also includes EMO_1 as an additional control variable (both specifications are estimated with a Probit model using IVs for EMO_1 and EMO_{t-1}). In both specifications, the lagged *EMO* has a positive effect on the probability of finishing the conversation. Hence, positive customer emotions are an indication that the conversation is closer to completion. One interpretation of this result is that customer emotion is a proxy for case complexity, where less complex cases–which take fewer turns to complete–are presented by customers expressing positive emotion. Including EMO_1 as a covariate rules out this explanation: the emotion in the first customer message controls for the initial emotion of the customer which can be related to the case complexity. We observe that the effect of lagged emotion EMO_{t-1} is very similar when including or excluding EMO_1 as a control variable, suggesting that the effect is not confounded by unobserved factors related to case complexity. The effect is also economically significant: changing EMO_{t-1} from -1 (negative) to 0 (neutral) increases the probability of ending the conversation from 0.08 to 0.2, on average.

Altogether, the results suggest direct and indirect paths through which customer emotion affects agent behavior. First, the agent spends more effort writing to customers with negative emotion (compared to neutral emotion), which increases *RT*. But this mechanism explains only a small fraction of the increase in *RT*. For agent messages of similar length, the results suggest that *RT* continues to be longer for customers with negative emotion relative to neutral and positive emotion. Moreover, conversations that start with more negative customer emotion tend to be longer. This effect persists through the conversation: in any turn during the conversation, the remaining extension of the case increases when the customer is expressing negative emotion.

Table 5 - Effect of Customer Emotion on the length of a conversation (Outliers Excluded, Both EMO_{t-1} and EMO_1 are Instrumented using CustSent)

	Model (4) $Nturns$	Model (5) $Pr(LastTurn)$	Model (5) $Pr(LastTurn)$
EMO_1	-1.684***		-0.033***
	(0.0689)		(0.0069)
EMO_{t-1}		0.571***	0.577***
		(0.0042)	(0.0044)
$Concurrent$ (chat level)	-1.237***		
	(0.0369)		
$Concurrent_{t-1}$		0.019***	0.019***
		(0.003)	(0.003)
$NumInQueue$ (chat level)	0.032***		
	(0.0047)		
$NumInQueue_{t-1}$		0.001	0.001
		(0.0005)	(0.0005)
$log(CustWords_1)$	-0.328***		
	(0.0211)		
$log(CustWords_{t-1})$		-0.016***	-0.015***
		(0.0018)	(0.0018)
$Turn_t$		-0.005***	-0.005***
		(0.0005)	(0.0005)
$IsWeekend$	-0.012	0.008	0.008
	(0.0415)	(0.0045)	(0.0045)
$SrvType$	6.192*	-0.277*	-0.278*
	(3.1458)	(0.1274)	(0.1275)
$ShiftTime$	Included	Included	Included
$HourOfDay$	Included	Included	Included
Agent Fixed Effect	Included	Included	Included
Constant	11.129***	-0.886***	-0.883***
	(1.6750)	(0.1454)	(0.1454)
Observations	141,654	518,437	518,437

Standard errors in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

4.4.2 Effect of Agent Behavior on Customer Emotions

Table 6 shows the results of the models with customer emotion *EMO* as the dependent variable (Models (6) and (7)). Recall this specification uses each message as a unit of analysis and includes fixed effects for the conversation, so the identification is based on variation across turns within a conversation. Two specifications were estimated, including different sets of covariates that measure distinct aspects of agent behavior. The specification reported in the first column corresponds to Model (6), which includes *RT* as the main covariate. Recall that this estimation is carried out using exogenous workload-related IVs (*Concurrent* and *NumInQueue*) instrumenting *RT*, in order to remove the variation in *RT* that could be driven by the unobserved quality of the agent response (we discusses results of the estimation without IVs in the Sensitivity analysis below). The estimation suggests that doubling *RT* decreases customer emotion by 0.06, equivalent to less than 0.1 standard deviations, a relatively small effect.

Table 6 - Effect of Agent Behavior on Customer Emotion (Outliers Excluded. log(RT_{t-1}) is Instrumented using $Concurrent_{t-1}$ and $NumInQueue_{t-1}$)

	Model (6) EMO	Model (7) EMO
$log(RT_{t-1})$	-0.062***	-0.427***
	(0.0145)	(0.0403)
$ConvStage_t$	0.896***	1.181***
	(0.0057)	(0.0079)
$log(NumWords_{t-1})$		0.200***
		(0.0185)
$Turn_t$		-0.016***
		(0.0003)
Conversation Fixed Effect	Included	Included
Constant	0.066	0.794***
	(0.0524)	(0.0943)
Observations	586,456	586,456

Standard errors in parentheses
$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

The second specification corresponds to Model (7), including *RT*, log(*NumWords*) and the corresponding *Turn* number as the main covariates and using the same IVs as in the previous specification to instrument *RT*. The results reveal that customer emotion becomes more positive for longer messages: doubling *NumWords* increases *EMO* by 0.2. One interpretation of this result is that customers find longer agent messages to be more informative or a signal that the agent is paying attention to them, thereby improving their emotion. In addition, customer emotion tends to decrease for longer conversations: an increase by 10 turns (equal to the average number of turns in a conversation) reduces customer emotion by 0.16. Furthermore, controlling for these other measures of agent behavior reveals a larger effect of *RT* on customer emotion: doubling response time decreases *EMO* by 0.43, which is about half the standard deviation of the dependent variable.

Overall, the results suggest that customer emotion is affected by the different measures of agent performance-related behaviors, where the predominant effect is a negative effect of *RT* on customer emotion. The managerial implications of these results are discussed in Section 4.5.

4.4.3 Sensitivity Analysis and Alternative Specifications

We analyzed several alternative specifications of the models to verify the robustness of the empirical results, which are summarized in this section. All the result tables of these additional analyses are reported in Appendix 5-17.

Models (1)–(4) in Table 4 and Table 5 are estimated with 2SLS instrumenting *EMO* with an alternative sentiment measure, in order to mitigate attenuation bias due to measurement error. For robustness, the same specifications were estimated with Ordinary Least Squares (OLS), without instruments (Appendix 5). The results reveal a negative effect of *EMO* on *RT*, which is statistically significant and smaller in magnitude compared to the estimations reported in Table 4 (coefficient is approx. -0.1 compared to -0.2). In Model (3) the effect of *EMO* on *NumWords* is smaller in magnitude and not statistically significant. For Model (4), with *NTurns* as dependent variable, the coefficient of EMO_1 changes from -1.684 to -0.250 (p-value < 0.001). Overall, these results are consistent with attenuation bias due to imprecise measurement of customer emotion, which can be corrected with the IV estimation proposed using an alternative sentiment measure as instrument.

Our analysis uses *SentiStrength* as the main measure of customer emotion and uses *CustSent* as an IV to correct for measurement error. An alternative approach is to combine both measures into one variable using factor analysis, where the first factor is used as a construct for customer emotion. Doing so yields similar results to those reported in Table 4 in terms of magnitude and statistical significance (Appendix 6). One difference is that when factor analysis is used to measure the effect of emotion on number of words (Equation (3)) the effect is smaller in magnitude and significant at the 0.1 level. Consequently, the indirect effect of *EMO* on *RT* (via the number of words) is smaller in this specification, about half the magnitude compared to the original results. Models (1)–(4) were also estimated replacing *EMO* with the alternative *CustSent* measure (see Appendix 17). The results are similar to the main results presented above.

In addition, we estimated Models (1)–(3) without log transformation to the dependent variable, and Model (4) with log transformation to the dependent variable. The results were similar in terms of the signs, magnitude and statistical significance (Appendix 7).

Model (2) includes the number of words (*NumWords*), a proxy of agent effort, as a mediator. Another possible mediation is the emotion expressed in the agent's message, which we measured using the same sentiment analysis tools (see footnote 1). To check the robustness of our results, we included a second mediator–emotion in the agent's message. Including this variable in our analyses did not change our main results: the effect of *EMO* and *NumWords* remained similar to those reported in Table 4 (see Appendix 8), with a slightly smaller coefficient for *EMO* (drops from 0.2 to 0.16). The results suggest that the emotion in the agent's message is negatively related to *RT*. Emotions expressed by agents are highly influenced by organizational requirements regarding appropriate emotional displays. Additionally, agent expressed emotion is endogenous to agent RT, making it difficult to infer causality. Therefore, this analysis requires further investigation where agent expressed emotion is the main focus, and only presented here as a robustness test.

All the models related to agent behavior include *Concurrent* as a control variable to account for the effects of multitasking. Our measure of concurrency is calculated based on number of simultaneous conversations assigned to an agent during the focal conversation. For robustness, we also estimated the models using two alternative definitions of concurrency; (i) the number of words and (ii) the number of messages written by the agent in parallel conversations. In all cases, concurrency has a positive effect on *RT*, corroborating that multitasking indeed increases the *RT* in a focal conversation (Appendix 9). The effect of *EMO* on *RT* is similar to the main results (reported in Table 4) across all the specifications with alternative measures of concurrency. For our main analysis we preferred using the number of simultaneous conversations as a measure of concurrency because this is exogenous to the agent, whereas the number of words (or messages) written in parallel is endogenous.

The estimation of Model (1) is carried out using IVs and panel data, including fixed effects and assuming i.i.d. (independent and identically distributed) random errors. Examining the residuals of the model reveals serial correlation, and therefore the calculation of the standard errors may not be accurate. We estimated the same model clustering observations at the conversation level, which allows for arbitrary correlation within clusters. The standard errors were similar and the main conclusions do not change (Appendix 10). In Table 4, the estimates of Models (2) and (3) use bootstrapped standard errors which account for correlation between the error terms within cluster.

Table 5 includes customer emotion as a linear predictor of the number of turns (Models (4) and (5)). For robustness, we also estimated the models including *EMO* in three levels, capturing positive, neutral and negative emotion (Appendix 11). The results for Model (4) reveal a monotone non-linear effect of the emotion of the first message: taking neutral emotion as the base, positive emotion reduces *NTurns* by 0.726 whereas negative emotion leads to an increase of 3.463 turns. In the survival Model (5), the coefficient associated to positive emotion in the previous message is positive and much larger in magnitude relative to negative emotion, consistent with a positive effect of emotion on the likelihood of ending the conversation (and thereby a shorter length of the conversation). The main conclusions remain when using a non-linear specification for the effect of customer emotion.

Recall that the results from Table 4 and Table 5 are based on a sample without outliers. The same models were estimated with all the observations, including the outliers (Appendix 12-13). Overall, the conclusions obtained from these results are similar. The coefficients associated to *EMO* in Models (1) and (2) continue to be negative and similar in magnitude. For Model (3), the effect of *EMO* on *NumWords* is negative and significant, with a point estimate of -0.024. As discussed previously, the additional analysis reported in Table 4 showed that the effect of *EMO* on *NumWords* appears to be non-linear, suggesting that the model with linear *EMO* is not well specified and less robust. This may explain why the *EMO* coefficient in the linear specification is sensitive to the definition of the sample (Appendix 12). In Model (4), with *NTurns* as dependent variable, the coefficient on EMO_1 , -1.677, is similar to the main analysis (Table 5).

Additional analysis was carried out to evaluate the robustness of Models (6) and (7) (main results reported in Table 6), with customer emotion used as the dependent variable. Recall that these models are estimated with IVs to address the endogeneity of *RT*, which is potentially correlated with unobservable factors associated to the quality of the response. The same models were estimated without IVs using OLS (Appendix 15). The coefficient associated with log(*RT*) flips from negative to positive, with a point estimate close to 0.02 (with p-value < 0.001). This is consistent with the endogeneity bias that was conjectured: because *RT* is likely to be positively correlated with quality, which is unobservable, and is part of the error term. This generates a positive bias in the estimated coefficient of *RT*. Instrumenting *RT* with exogenous factors associated to agent workload helps to correct this bias. Models (6) and (7) were also estimated using OLS regression and replacing *EMO* with the alternative *CustSent* measure (Appendix 16). The results are similar to the main analysis.

Overall, the sensitivity analysis provides further support of the estimation results, showing that they are robust to alternative specifications.

4.5. Managerial Implications

The fact that customer emotion impacts agent behavior and that agent behavior impacts customer emotion suggests that emotional workload should be monitored and taken into account in operational decisions. This section discusses some prescriptions for the design and control of service systems that are subject to the effects of emotional workload. This is even more important considering findings that suggest that customer emotions reflect customer satisfaction (Ashtar, Rafaeli, & Yom-Tov, 2020; Yom-Tov et al., 2018) and the connection of the latter to organization profitability.

4.5.1 Performance goals, system design and staffing

The results presented here should serve as a "call for awareness" that emotional workload exists, varies within a service encounter, and impacts agent performance. A standard approach is to consider service time and case characteristics as the key dimensions of load, but our work suggests that customer emotion is another important factor. Dealing with more negative customers will require agents to spend more time to solve the customer issues and to cope with customers emotion. To evaluate the total effect of emotion, it is useful to compare the agent time required to handle angry (negative), neutral, and happy (positive) customers. To do that we define total throughput time by multiplying the average agent RT per turn by the average number of turns in a chat. The total throughput time required to handle a “negative” customer is 15.7 minutes (1.32 minutes times 11.9 turns); compared to 11.1 minutes for a neutral customer and 7.6 minutes to handle a “positive” one. This analysis suggests that the total amount of throughput time associated with a negative customer is 42% longer than the one associated with a neutral customer. Many contact centers measure agent performance by the number of calls an agent handles per hour (average concurrency divided by total throughput time). They should be aware that an agent can serve 12.6 neutral customers per hour but only 8.9 negative ones (assuming average concurrency=2.33). Hence, the evaluation of service agents or teams who encounter a high proportion of negative customers (for example customer retention teams) should be based on adjusted targets of calls per hour. This is important if a contact center is considering a design change to incorporate skill-based routing (i.e. that each customer group is served by a separated agent-skill group).

Another way to think about the implication of emotional workload is to consider how variations in customer mix impact the offered load, which is equivalent to the amount of staffing required to handle the arriving customer workload. We present in Table 7 a comparison between the offered load (calculated by the arrival rate, given in Appendix 18, multiplied by total throughput times divided by average concurrency) of the current mix of emotional messages in the contact center we analyzed vs. the offered load that the agents will need to handle if 10% of the messages transformed from being neutral to negative for some reason. Such a situation might arise after a company experiences failure in one of its products or services. This kind of change in the mix of customer emotion would increase the number of agents needed to handle customer issues by 4.4% and the amount of text written per day by 2.2%, assuming no change in arrival rate. This analysis suggest that customer emotion is an important factor that should be accounted for in staffing decisions.

4.5.2 Counterfactual analysis: the impact of emotional workload on system-level performance

Most organizations do not monitor customer emotions and do not adjust staffing to match variation in customer emotion. Here we would like to calculate the impact of an increase in emotional workload on performance level, when staffing remains constant. We use the same scenario presented in Table 7.

Table 7 - Comparison of Working Hour Associated with Different Mix of Message Emotion

Emotion Type	Base Case			Alternative Case		
	Message percentage	Offered Load	Effort (NumWords)	Message percentage	Offered Load	Effort (NumWords)
Positive	29%	14.8	63,685	29%	14.8	63,685
Neutral	63%	47.1	162,652	53%	39.6	136,834
Negative	8%	8.5	25,103	18%	19.1	56,483
Total	100%	70.4	251,440	100%	73.7	257,002

We use a simulation model that was calibrated to the operation of contact centers and developed by Castellanos et al. (2019). It is a version of an Erlang-A model that takes into account unique contact center features such as silent abandonment. For this counterfactual analysis, we estimated the simulation parameters using the data from February 2017. The simulation assumes that customers arrive to the contact center according to a non-homogeneous Poisson process with rate $\lambda_{d,t}$ ($\lambda_{d,t}$ is the arrival rate at day d and hour t). The customers are served by $n_{d,t}$ statistically identical agents, with average concurrency of c. Therefore, the number of service slots available at time (d,t) is $cn_{d,t}$. If there is no available slot, the customer waits in a First-Come-First-Serve queue. The customer has finite patience, assumed to be exponentially distributed with rate θ, which was estimated using the methodology developed in Yefenof et al. (2018) and that takes into account the fact that customer waiting time in chat systems is censored both from the right (by the offered wait) and from the left (by the fact that sometimes customers abandon the service without exiting the queue–they do not close the chat window but "disappear"). For more details about this silent abandonment phenomenon and its implication see Castellanos et al. (2019). In our data θ=0.5, and

70% of the customers indicate their abandonment in real time (30% abandon silently). Service times are assumed to be exponentially distributed with rate μ. Note that μ in this simulation is 1 divided by the throughput time of a conversation and equal to μ=0.075; in the counterfactual scenario with more negative customers, μ was adjusted so that the throughput time was 4.4% longer (as suggested by our empirical results).

The simulation predicts that the 10% change from neutral to negative messages increases the probability of abandonment by 2% and increases the expected waiting time by two minutes.

4.5.3 **Routing policies designed to achieve load balancing or specialization**

Emotional workload should also impact work allocation (routing) decisions. In our data, agents usually serve up to three (average 2.3) customers simultaneously. However, the load created by three negative customers differs dramatically from the load created by three positive customers. Specifically, dealing with three neutral customer messages is equivalent (in terms of workload) to dealing with only 2.5 negative messages or 3.7 positive messages. We suggest that like other measures of workload, emotional workload could be used in the design of dispatching rules commonly used in contact centers to *dynamically* adjust the workload of agents based on real-time assessments. The sentiment analysis tool used in this work allows for real-time monitoring of emotional workload during service conversations. Previous research showed that there is a clear trend of sentiment within a conversation (Yom-Tov et al., 2018): customer sentiment usually starts negative, then becomes neutral and transforms to positive toward the end of the conversation. This positive trend is captured in our analysis by the variable *ConvStage* that monitors the conversation progress.

We suggest designing a routing policy that would *balance* both offered load and emotional workload. The idea is that when a new conversation arrives it will be assigned to an agent that has the least overall load including offered, and emotional. Such a policy will dynamically allocate more capacity to agents that handle customers who consistently express negative emotion. This dynamic allocation can be based on Model (5), and would allow an agent to spend more time dealing with a negative customer which would be expected to improve customer emotion and overall customer satisfaction. This idea draws its intuition from (Armony & Ward, 2010) and Mandelbaum et al. (Mandelbaum, Momčilović, & Tseytlin, 2012), who suggest adoption of allocating policy that is fair from the agent perspective (Carmeli, Yom-Tov, & Mandelbaum, 2018).

In addition, in some contact centers customers write their inquiry before entering the queue (Castellanos et al., 2019). In such cases, we can assess the emotional workload expected by that customer inquiry in real-time before assigning a new conversation to an agent and using the measure of emotional workload we can also predict that customer needs. Model (4) supports the claim that this is indeed possible by showing that one can predict the number of turns that a chat will require using the customer sentiment of the first turn. This information can be used for designing a *"sentiment based routing"* mechanism, analogous to skill-based routing. This routing mechanism could also assign an emotional call to the most appropriate agent group (e.g., customer retention team) trained to deal with particular customer emotion (e.g., anger).

4.5.4 **Prioritization**

Our results show that longer agent RT hampers customer emotions. Therefore, operational policies that reduce agent RT will improve customer emotions. Such policies might be implemented in the following way: since agents handle multiple customers in parallel, they might miss expressions of negative customer emotion while they are interacting with other customers. Real-time monitoring can increase agents awareness by alerting them when an escalation in negative emotion occurs. For example sentiment engines can be designed to provide real-time monitoring of customer sentiment, and alert managers and agents of problematic situations (e.g., when the sentiment of a customer drops below a specific threshold). These alerts will enable agents to *prioritize* unsatisfied customers, reduce their RT, and improve customer satisfaction. Moreover, managers can use these alerts to identify extreme negative sentiment cases, and to provide agents with relevant assistance. This idea is now being implemented in some of the companies working with the LivePerson platform. In addition, given our finding that longer agent texts improve customer sentiment, agents should be made aware of the impact of message length and provided with indication of when customer sentiment is deteriorating, by alerts similar to those suggested above. In these cases agents should be trained to react by communicating their effort better in order to improve customer emotion.

5. Discussion

Our findings show that customer expressed emotion, a form of emotional demand, impacts agent performance-related behaviors: agents respond more slowly and write more words to customers who express negative emotion, compared to positive or neutral emotion, supporting Hypotheses 1 and 2, respectively. Negative customer emotion increases agent RT directly and indirectly through agent effort (supporting Hypothesis 3). Most of the effect of customer emotion, however, is direct (see Table 4). This suggests that there may be additional mechanisms through which customer emotion impacts agent RT. One option is that our proxy of effort captures only a portion of agent effort. Future research should include data about other agent activities to fully understand the role of agent effort. Another option is that agents prioritize customers depending on their expressed emotions. For example, recent findings suggest that decision-makers' perceptions impact patient prioritization in Emergency Departments (Ding, Park, Nagarajan, & Grafstein, 2019). We call for future research to continue this line of work to understand how service agents prioritize concurrent customers and whether emotional workload impacts prioritization.

In addition, we showed that negative customer emotion prolongs the service interaction, supporting Hypothesis 4. This effect is large, and one possible mechanism may be agent errors

(Rafaeli et al., 2012): when agents encounter expressions of negative emotion, they are more likely to make mistakes, extending the service encounter as a result. We cannot test this mechanism in the current dataset because we cannot automatically code agent errors in the data. We hope that future advancements in the field of Natural Language Processing will help researchers in pursuing this direction.

Overall, our findings suggest that negative customer emotions create a burden on agents, and that positive customer emotions may act as a source of motivation (Bakker & Demerouti, 2007). The estimations suggest that the effects of emotional workload are substantial, and comparable to workload factors. For example, a negative customer message increases agent RT in the subsequent message by 4.2% similar to adding another customer to an agent which increases RT by 7.4% (see Table 4). Our findings also show that the opposite direction of influence exists– an increase in agent RT or an increase in the number of turns, hampers customer emotion (supporting Hypotheses 5 and 7, respectively). An increase in agent effort, however, leads to greater positive customer emotion as predicted (Hypothesis 6).

Previous research relied almost solely on experimental manipulations with small samples and low-resolution self-reported emotions, thus affording limited managerial insights. Study 3 overcomes these limitations by using operational and objective measures of agent behavior and of customer emotion in real service conversations, measured at the resolution of a single message. We show that emotional workload creates “micro-level influences,” that occur at the level of a single message within the conversation between an agent and a customer. We theorize and show empirical effects of emotional workload that goes beyond multitasking and queue length effects. Our analyses of a large data-set of conversations between agents and customers, empirically measure this type of load, and document its influence on critical Operations Management parameters including agent RT, agent effort, and the number of turns it takes to complete a conversation.

We introduce a new position for customer emotion in service–that of a potential source of load. This is in contrast to traditional Operations Management views, where customer emotion is treated as an outcome. The implicit assumption of past research was that customer happiness depends on their evaluation of the quality of service. We show, however, that customer emotions may also be a factor that determines the efficiency of the service. This suggests that the concept of load actually comprises multiple aspects, and that emotional workload is one of them. This view of load accounts for pressures beyond the mere presence of a customer, and is based on factors inherent to the nature and content of individual service conversations.

Emotions provide information (data) about a social situation and the actors in it (van Kleef, 2015). To date, these data have served only the service dyad: an agent and a customer. This dyad is engaged in co-production of value; both actors invest effort to resolve a specific issue. The ratio of the effort between the service interaction partners is dependent on context. For example, if a customer requests easy-to-get information, the ratio of effort will be close to 1. In contrast, if a customer has a complicated request, or if the customer creates high emotional workload, the agent will likely need to invest more effort than the customer. As Roels (2014) showed, one can improve service system efficiency by considering the effort ratio and route customers to adequate service channels based on it. We therefore call for researchers and practitioners to view customer emotion as data that can aid them in designing service systems.

The type of data we use in Study 3 is increasingly available in service organizations (i.e., full documentation of service). We highlight the opportunities that such data, coupled with automated sentiment analysis tools create for studying service delivery (Rafaeli et al., 2017, 2019) and improving the operations of contact centers. From a managerial perspective, our analyses suggest the importance of incorporating real time monitoring of the emotions of customers being handled by service agents. Beyond the technical count of the number of customers in the service system, service operations need to acknowledge the dynamics that customers bring to the system. This includes the types of problems that customers raise, the verbosity with which they communicate their problems, and the emotions that they attach to this communication. Failing to recognize such customer-induced states can lead to inaccurate planning models and sub-optimal service management.

References

Allon, G., Federgruen, A., & Pierson, M. (2011). How much is a reduction of your customers' wait worth? An empirical study of the fast-food drive-thru industry based on structural estimation methods. *Manufacturing & Service Operations Management*, *13*(4), 489–507.

Amazon Web Services, I. (2021). Amazon Rekognition Automate your image and video analysis with machine learning. Retrieved from aws.amazon.com website: https://aws.amazon.com/rekognition/

Anand, K. S., Paç, M. F., & Veeraraghavan, S. (2011). Quality–speed conundrum: Trade-offs in customer-intensive services. *Management Science, 57*(1), 40–56.

Armony, M., & Ward, A. R. (2010). Fair dynamic routing in large-scale heterogeneous-server systems. *Operations Research*, *58*(3), 624–637.

Ashtar, S., Rafaeli, A., & Yom-Tov, G. B. (2020). High-peak, low-peak, end and trend: The influence of features of emotions expressed by customers and employees on customer evaluations of frontline service. *Working Paper*.

Bakker, A. B., & Demerouti, E. (2007). The Job Demands-Resources model: state of the art. *Journal of Managerial Psychology*, *22*(3), 309–328.

Bakker, A. B., Demerouti, E., & Schaufeli, W. (2003). Dual processes at work in a call centre: An application of the job demands – resources model. *European Journal of Work and Organizational Psychology*, *12*(4), 393–417.

Beal, D. J., Weiss, H. M., Barros, E., & MacDermid, S. M. (2005). An episodic process model of affective influences on performance. *The Journal of Applied Psychology*, *90*(6), 1054–1068.

Bray, R. L., Coviello, D., Ichino, A., & Persico, N. (2016). Multitasking, multiarmed bandits, and the Italian judiciary. *Manufacturing & Service Operations Management*, *18*(4), 545–558.

Brief, A. P., & George, J. M. (1995). Psychological stress and the workplace: A brief comment on Lazarus outlook. In R. Crandall & P. L. Perrewe (Eds.), *Occupational stress: a handbook.* Philadelphia: Taylor and Francis.

Butterfield, L. D., Borgen, W. A., Amundson, N. E., & Maglio, A.-S. T. (2005). Fifty years of the critical incident technique: 1954-2004 and beyond. *Qualitative Research*, *5*(4), 475–497.

Carayon, P., & Alvarado, C. J. (2007). Workload and Patient Safety Among Critical Care Nurses. *Critical Care Nursing Clinics of North America*, *19*(2), 121–129.

Carmeli, N., Yom-Tov, G. B., & Mandelbaum, A. (2018). *Multi-dimensional load: Balancing emotional and system load in service systems*. Working paper.

Casado Díaz, A. B., & Más Ruíz, F. J. (2002). The consumer's reaction to delays in service. *International Journal of Service Industry Management*, *13*(2), 118–140.

Castellanos, A., Yom-Tov, G. B., & Goldberg, Y. (2019). *Silent abandonment in contact centers: Estimating customer patience with uncertain data*. Working paper.

Cho, D. D., Bretthauer, K. M., Cattani, K. D., & Mills, A. F. (2019). Behavior Aware Service Staffing. *Production and Operations Management*, *28*(5), 1285–1304.

Crawford, E. R., LePine, J. A., & Rich, B. L. (2010). Linking job demands and resources to employee engagement and burnout: A theoretical extension and meta-analytic test. *Journal of Applied Psychology*, *95*(5), 834–848.

Czerwinski, M., Horvitz, E., & Wilhite, S. (2004). A diary study of task switching and interruptions. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 175–182.

Dai, H., Milkman, K. L., Hofmann, D. A., & Staats, B. R. (2015). The impact of time at work and time off from work on rule compliance: the case of hand hygiene in health care. *Journal of Applied Psychology*, *100*(3), 846.

Dasgupta, P. (2012). Effect of Role Ambiguity, Conflict and Overload in Private Hospitals'

Nurses' Burnout and Mediation Through Self Efficacy. *Journal of Health Management*, *14*(4), 513–534.

De Vaus, J., Hornsey, M. J., Kuppens, P., & Bastian, B. (2018). Exploring the East-West Divide in Prevalence of Affective Disorder: A Case for Cultural Differences in Coping With Negative Emotion. *Personality and Social Psychology Review*, *22*(3), 285–304.

Delasay, M., Ingolfsson, A., Kolfal, B., & Schultz, K. (2019). Load effect on service times. *European Journal of Operational Research*, *279*(3), 673–686.

Ding, Y., Park, E., Nagarajan, M., & Grafstein, E. (2019). Patient prioritization in emergency department triage systems: An empirical study of the Canadian triage and acuity scale (CTAS). *Manufacturing & Service Operations Management*, *21*(4), 723–741.

Donabedian, A. (1988). The Quality of Care: How Can It Be Assessed? *JAMA: The Journal of the American Medical Association*, *260*(12), 1743–1748.

Donaldson, S. I., & Grant-vallone, E. J. (2002). Understanding Self-Report Bias in Organizational Behavior Research. *Journal of Business and Psychology*, *17*(2), 245–260.

Donoso, L. M. B., Demerouti, E., Garrosa Hernández, E., Moreno-Jiménez, B., & Carmona Cobo, I. (2015). Positive benefits of caring on nurses' motivation and well-being: A diary study about the role of emotional regulation abilities at work. *International Journal of Nursing Studies*, *52*(4), 804–816.

Drach-Zahavy, A., Yagil, D., & Cohen, I. (2017). Social model of emotional labour and client satisfaction: exploring inter- and intrapersonal characteristics of the client−provider encounter. *Work and Stress*, *31*(2), 182–208.

Felton, J. S. (1998). Burnout as a clinical entity - its importance in health care workers. *Occupational Medicine*, *48*(4), 237–250.

Field, J. M., Victorino, L., Buell, R. W., Dixon, M. J., Goldstein, S. M., Menor, L. J., … Zhang, J. J. (2018). Service operations: what's next? *Journal of Service Management*, *29*(1), 39–54.

Fredrickson, B. L., & Kahneman, D. (1993). Duration neglect in retrospective evaluations of affective episodes. *Journal of Personality and Social Psychology*, *65*(1), 45.

Fritz, C., Lam, C. F., & Spreitzer, G. M. (2011). An Examination of Knowledge Workers' Energy Management. *Academy of Management Perspectives*, *25*(3), 28–40.

Gabriel, A. S., & Diefendorff, J. M. (2015). Emotional labor dynamics: A momentary approach. *Academy of Management Journal*, *58*(6), 1804–1825.

Geddes, D., & Callister, R. R. (2007). Crossing the line (s): A dual threshold model of anger in organizations. *Academy of Management Review*, *32*(3), 721–746.

Goes, P. B., Ilk, N., Lin, M., & Zhao, J. L. (2018). When More is Less: Field Evidence on Unintended Consequences of Multitasking. *Management Science*, *64*(7), 3033–3054.

Goldberg, L. S., & Grandey, A. A. (2007). Display rules versus display autonomy: emotion regulation, emotional exhaustion, and task performance in a call center simulation. *Journal of Occupational Health Psychology*, *12*(3), 301.

Grandey, A. A., Dickter, D., & Sin, H. P. (2004). The customer is not always right: Customer verbal aggression toward service employees. *Journal of Organizational Behavior*, *25*(3), 397–418.

Grandey, A. A., Rafaeli, A., Ravid, S., Wirtz, J., & Steiner, D. D. (2010). Emotion display rules at work in the global service economy: the special case of the customer. *Journal of Service Management*, *21*(3), 388–412.

Groth, M., Hennig-Thurau, T., & Walsh, G. (2009). Customer reactions to emotional labor: The roles of employee acting strategies and customer detection accuracy. *Academy of Management Journal*, *52*(5), 958–974.

Guest, G., Bunce, A., & Johnson, L. (2006). How Many Interviews Are Enough?: An Experiment with Data Saturation and Variability. *Field Methods*, *18*(1), 59–82.

Hall, R. W. (1991). *Queueing methods: for services and manufacturing*. Pearson College Division.

Hareli, S., & Rafaeli, A. (2008). Emotion cycles: On the social influence of emotion in organizations. *Research in Organizational Behavior*, *28*, 35–59.

Hayes, A. F., & Rockwood, N. J. (2020). Conditional process analysis: Concepts, computation, and advances in the modeling of the contingencies of mechanisms. *American Behavioral Scientist*, *64*(1), 19–54.

Jackson, S. E., Schwab, R. L., & Schuler, R. S. (1986). Toward an Understanding of the Burnout Phenomenon. *Journal of Applied Psychology*, *71*(4), 630–640.

James, L. R., Demaree, R. G., & Wolf, G. (1984). Estimating within-group interrater reliability with and without response bias. *Journal of Applied Psychology*, *69*(1), 85.

Katz, K., Larson, B., & Larson, R. (1991). Prescription for the waiting-in-line blues: Entertain, enlighten, and engage. *MIT Sloan Management Review*, *32*(2), 44–55.

Kc, D. S. (2014). Does Multitasking Improve Performance ? Evidence from the Emergency Department. *Manufacturing & Service Operations Management*, *16*(2), 168–183.

Kc, D. S., & Terwiesch, C. (2009). Impact of Workload on Service Time and Patient Safety: An Econometric Analysis of Hospital Operations. *Management Science*, *55*(9), 1486–1498.

Knutson, B. (1996). Facial expressions of emotion influence interpersonal trait inferences. *Journal of Nonverbal Behavior*, *20*(3), 165–182.

Kooij, J. F. P., Liem, M. C., Krijnders, J. D., Andringa, T. C., & Gavrila, D. M. (2016). Multi-modal human aggression detection. *Computer Vision and Image Understanding*, *144*, 106–120.

Landau, S. F., & Bendalak, Y. (2008). Personnel exposure to violence in hospital emergency wards: a routine activity approach. *Aggressive Behavior*, *34*(1), 88–103.

Larson, R. C. (1987). OR forum—perspectives on queues: Social justice and the psychology of queueing. *Operations Research*, *35*(6), 895–905.

Lewis, P. S., & Malecha, A. (2011). The impact of workplace incivility on the work environment, manager skill, and productivity. *Journal of Nursing Administration*, *41*(1), 41–47.

Llorens, S., Bakker, A. B., Schaufeli, W., & Salanova, M. (2006). Testing the robustness of the job demands-resources model. *International Journal of Stress Management*, *13*(3), 378–391.

Maister, D. (1985). The psychology of waiting lines. In J. D. Czepiel, M. R. Solomon, & C. F. Surprenanat (Eds.), *The Service Encounter* (pp. 24–113). Lexington, MA: Lexington Books.

Mandelbaum, A., Momčilović, P., & Tseytlin, Y. (2012). On Fair Routing from Emergency Departments to Hospital Wards: QED Queues with Heterogeneous Servers. *Management Science*, *58*(7), 1273–1291.

Mandelbaum, A., & Zeltyn, S. (2013). Data-stories about (im) patient customers in tele-queues. *Queueing Systems*, *75*(2–4), 115–146.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, *60*(3), 531–542.

Miron-Shatz, T. (2009). Evaluating Multiepisode Events: Boundary Conditions for the Peak-End Rule. *Emotion*, *9*(2), 206–213.

Mohr, L. A., & Bitner, M. J. (1995). The role of employee effort in satisfaction with service transactions. *Journal of Business Research*, *32*(3), 239–252.

Motowidlo, S. J., Packard, J. S., & Manning, M. R. (1986). Occupational Stress. Its Causes and Consequences for Job Performance. *Journal of Applied Psychology*, *71*(4), 618–629.

Ohly, S., & Schmitt, A. (2015). What Makes Us Enthusiastic, Angry, Feeling at Rest or Worried? Development and Validation of an Affective Work Events Taxonomy Using Concept Mapping Methodology. *Journal of Business and Psychology*, *30*(1), 15–35.

Ohly, S., & Venz, L. (2021). *Emotion and proactivity at work: Where are we now?* (K. Z. P. & C.-H. Wu, Ed.). The Bristol University Press.

Penrod, J., Preston, D. B., Cain, R. E., & Starks, M. T. (2003). A Discussion of Chain Referral As a Method of Sampling Hard-to-Reach Populations. *Journal of Transcultural Nursing*, *14*(2), 100–107.

Peräkylä, A., Henttonen, P., Voutilainen, L., Kahri, M., Stevanovic, M., Sams, M., & Ravaja, N. (2015). Sharing the Emotional Load: Recipient Affiliation Calms Down the Storyteller. *Social Psychology Quarterly*, *78*(4), 301–323.

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies. *Journal of Applied Psychology*, *88*(5), 879–903.

Porath, C. L., & Erez, A. (2007). Does rudeness really matter? The effects of rudeness on task performance and helpfulness. *Academy of Management Journal*, *50*(5), 1181–1197.

Pradhan, S., & Jena, L. K. (2018). Emotional intelligence as a moderator in abusive supervision-intention to quit relationship among Indian healthcare professionals. *Asia-Pacific Journal of Business Administration*, *10*(1), 35–49.

Rafaeli, A., Altman, D., Gremler, D. D., Huang, M. H., Grewal, D., Iyer, B., … Ruyter, K. (2017). The Future of Frontline Research: Invited Commentaries. *Journal of Service Research*, *20*(1), 91–99.

Rafaeli, A., Ashtar, S., & Altman, D. (2019). Digital Traces: New Data, Resources, and Tools for Psychological-Science Research. *Current Directions in Psychological Science*, *28*(6), 560–566.

Rafaeli, A., Erez, A., Ravid, S., Derfler-Rozin, R., Treister, D. E., & Scheyer, R. (2012). When customers exhibit verbal aggression, employees pay cognitive costs. *Journal of Applied Psychology*, *97*(5), 931.

Rafaeli, A., & Sutton, R. I. (1987). Expression of emotion as part of the work role. *Academy of Management Review*, *12*(1), 23–37.

Rafaeli, A., & Sutton, R. I. (1989). The expression of emotion in organizational life. *Research in Organizational Behavior*, *11*(1), 1–42.

Roels, G. (2014). Optimal Design of Coproductive Services: Interaction and Work Allocation. *Manufacturing & Service Operations Management*, *16*(4), 578–594.

Rothmann, S., Mostert, K., & Strydom, M. (2006). A psychometric evaluation of the job demand-resources scale in South Africa. *SA Journal of Industrial Psychology*, *32*(4), 76–86.

Russo, M., Buonocore, F., & Ferrara, M. (2015). Motivational mechanisms influencing error reporting among nurses. *Journal of Managerial Psychology*, *30*(2), 118–132.

Schaufeli, W. B., & Bakker, A. B. (2004). Job demands, job resources, and their relationship with burnout and engagement: A multi-sample study. *Journal of Organizational Behavior*, *25*(3), 293–315.

Schilpzand, P., De Pater, I. E., & Erez, A. (2016). Workplace incivility: A review of the literature and agenda for future research. *Journal of Organizational Behavior*, *37*, S57–S88.

Socher, R., Perelygin, A., & Wu, J. (2013). Recursive deep models for semantic compositionality over a sentiment treebank. *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, *1631*, 1642.

Song, H., Tucker, A. L., & Murrell, K. L. (2015). The diseconomies of queue pooling: An empirical investigation of emergency department length of stay. *Management Sience*, *61*(12), 3032–3053.

Sutton, R. I., & Rafaeli, A. (1988). Untangling the Relationships Between Displayed Emotions and Organizational Sales: the Case of Convenience Stores. *Academy of Management Journal*, *31*(3), 461–487.

Tan, T., & Netessine, S. (2014). When Does the Devil Make Work? An Empirical Study of the Impact of Workload on Worker Productivity. *Management Science*, *60*(6), 1574–1593.

Tausczik, Y. R., & Pennebaker, J. W. (2010). The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology*, *29*(1), 24–54.

Taylor, S. (1994). Waiting for service: the relationship between delays and evaluations of service. *Journal of Marketing*, *58*(2), 56–69.

Thelwall, M. (2013). Heart and soul: Sentiment strength detection in the social web with sentistrength. *Proceedings of the CyberEmotions*, *5*, 1–14.

Tiedens, L. Z. (2001). Anger and advancement versus sadness and subjugation: the effect of negative emotion expressions on social status conferral. *Journal of Personality and Social Psychology*, *80*(1), 86.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*(2), 207–232.

Van Der Heijden, B., Mahoney, C. B., & Xu, Y. (2019). Impact of job demands and resources on nurses' burnout and occupational turnover intention towards an age-moderated mediation model for the nursing profession. *International Journal of Environmental Research and Public Health*, *16*(11), 1–22.

van Kleef, G. A. (2015). Social Emotions Life Regulate The Emotions as Social Information (EASI) Model. *Current Directions in Psychological Science*, *18*(3), 184–188.

van Kleef, G. A., De Dreu, C. K. W., & Manstead, A. S. R. (2004). The interpersonal effects of anger and happiness in negotiations. *Journal of Personality and Social Psychology*, *86*(1), 57–76.

Van Veldhoven, M., & Meijman, T. F. (1994). The measurement of psychosocial job demands with a questionnaire (VBBA). *Amsterdam: NIA.*

Voutilainen, L., Henttonen, P., Kahri, M., Ravaja, N., Sams, M., & Peräkylä, A. (2018). Empathy, challenge, and psychophysiological activation in therapist-client interaction. *Frontiers in Psychology*, *9*, 1–13.

Walker, D. D., van Jaarsveld, D. D., & Skarlicki, D. P. (2017). Sticks and stones can break my bones but words can also hurt me: The relationship between customer verbal aggression and employee incivility. *Journal of Applied Psychology*, *102*(2), 163–179.

Wang, J., & Zhou, Y.-P. (2018). Impact of Queue Configuration on Service Time: Evidence from a Supermarket. *Management Science*, *64*(7), 3055–3075.

Wang, M., Liao, H., Zhan, Y., & Shi, J. (2011). Daily Customer Mistreatment and Employee Sabotage Against Customers:Examining Emotion and Resource Perspectives. *Academy of Management Journal*, *54*(2), 312–334.

Wansbeek, T., & Meijer, E. (2003). Measurement Error and Latent Variables. In B. H. Baltagi (Ed.), *A Companion to Theoretical Econometrics*. Blackwell Publishing Ltd.

Weiss, H. M., & Beal, D. J. (2005). Reflections on Affective Events Theory. In N. M. Ashkanasy, W. J. Zerbe, & C. E. J. Härtel (Eds.), *The effect of affect in organizational settings* (pp. 1–21). Bingley: Emerald Group Publishing Ltd.

Weiss, H. M., & Cropanzano, R. (1996). Affective Events Theory: A theoretical discussion of the structure, causes and consequences of affective experiences at work. *Research in Organizational Behavior*, *18*(1), 1–74.

Wittels, P., Johannes, B., Enne, R., Kirsch, K., & Gunga, H. C. (2002). Voice monitoring to measure emotional load during short-term stress. *European Journal of Applied Physiology*, *87*(3), 278–282.

World Health Organization. (2020). *State of the world's nursing 2020: investing in education, jobs and leadership*. https://www.who.int/publications/i/item/9789240003279

Yefenof, J., Goldberg, Y., Wiler, J., Mandelbaum, A., & Ritov, Y. (2018). Self-reporting and screening: Data with current-status and censored observations. *Working Paper*.

Yom-Tov, G. B., Ashtar, S., Altman, D., Natapov, M., Barkay, N., Westphal, M., & Rafaeli, A. (2018). Customer Sentiment in Web-Based Service Interactions: Automated Analyses and New Insights. *WWW '18 Companion: The 2018 Web Conference Companion*. Lyon, France.

Zohar, D., Epstein, R., & Tzischinski, O. (2003). Effects of Energy Availability on Immediate and Delayed Emotional Reactions to Work Events. *Journal of Applied Psychology*, *88*(6), 1082–1093.

Appendix

Appendix 1–The full list of events identified in Study 1, and rated in Study 2

Two subject matter experts independently coded the events into categories that were not predetermined. The authors then discussed the categorization until reaching agreement. The coding resulted in 16 categories that are presented in Table 8.

Table 8 - Events identified in Study 1 that can pose an emotional demand, divided by Category. Each event is presented with the number of taggers (taggers), number of taggers who experienced the event (taggers_valid), agreement score on the six emotional demand items (rwg.j), the average of all emotional demand items (ED_score) and averages of each item (ED1-ED6)

Q	Category	Event	Taggers	Taggers_ valid	rwg.j	ED_score	ED1	ED2	ED3	ED4	ED5	ED6
1	Co-worker issues	A doctor expresses distrust towards you	9	3	0.94	6.56	6.67	6.67	6.67	6.00	6.33	7.00
2	Co-worker issues	An intern doctor expresses distrust towards you	9	1	NA	5.00	5.00	5.00	5.00	6.00	3.00	6.00
3	Co-worker issues	A staff member doesn't arrive on time for their shift where you are working together	3	3	0.79	3.67	3.67	4.67	3.00	3.33	3.00	4.33
4	Co-worker issues	You need to treat a person that another staff member refuses to treat	5	3	0.00	4.17	4.33	5.00	2.67	5.33	5.00	2.67
5	Co-worker issues	A patient is mistreated by another staff member	5	3	0.00	4.94	5.33	4.67	3.00	4.33	6.00	6.33
6	Co-worker issues	You feel responsible for another staff member's mistake	3	3	0.00	3.28	4.33	4.33	1.33	2.33	3.00	4.33
7	Co-worker issues	You experience a conflict with a staff member	4	3	0.00	4.44	5.00	4.33	4.33	4.33	3.33	5.33

Table 8 - Continued

8	Co-worker issues	A staff member intervenes in your tasks	4	3	0.88	3.44	3.67	4.67	2.67	3.67	3.00	3.00
9	Co-worker issues	A colleague criticizes your professional decision	3	3	0.00	5.72	6.00	6.00	5.67	6.00	5.00	5.67
10	Co-worker issues	Other staff members speak in a language you don't understand	3	3	0.87	1.56	1.00	2.67	1.00	1.67	2.00	1.00
11	Co-worker issues	You have no friends working with you in your shift	5	3	0.00	4.17	4.67	4.67	2.00	4.67	5.33	3.67
12	Co-worker issues	You feel lonely during a shift	5	3	0.00	3.78	4.33	6.00	3.00	2.67	1.67	5.00
13	Co-worker issues	You are shouted at by another staff member	4	3	0.67	6.61	6.67	7.00	7.00	7.00	5.33	6.67
14	Co-worker issues	A staff member undermines you in front of patients	3	3	0.79	5.72	6.33	7.00	6.33	6.00	2.33	6.33
15	Co-worker issues	A doctor does not agree with your diagnosis	7	4	0.52	4.25	3.75	4.75	4.25	3.75	5.25	3.75
16	Co-worker issues	You call a doctor who is in a different ward	3	3	0.87	1.61	1.33	1.33	1.33	1.33	3.00	1.33
17	Co-worker issues	A treatment you need to give to a patient is delayed because of other staff members	3	3	0.00	4.94	4.67	4.67	4.33	5.67	5.67	4.67
18	Co-worker issues	A staff member in your shift repeatedly checks herself before doing something	7	4	0.69	2.13	2.50	1.75	1.25	2.50	3.00	1.75

Table 8 - Continued

19	Co-worker issues	A staff member pushes you	9	2	0.00	3.58	4.00	3.50	4.50	4.00	2.00	3.50
20	Co-worker issues	The staff in your department work inefficiently	3	3	0.88	4.56	4.67	5.67	2.00	5.67	5.33	4.00
21	Co-worker issues	There is a miscommunication between you and another staff member	3	3	0.92	3.56	2.67	3.33	2.00	4.33	5.00	4.00
22	Co-worker issues	A staff member takes a break when you don't have time for a break	4	4	0.00	3.33	3.25	4.00	2.75	4.00	3.25	2.75
23	Death	People crowd around a dying patient	4	3	0.00	4.83	6.00	5.67	2.33	5.33	4.00	5.67
24	Death	Your patient dies unexpectedly	4	3	0.00	3.61	5.33	3.67	2.00	4.33	2.00	4.33
25	Death	Your patient dies as a result of unprofessional treatment provided by you	5	1	NA	5.67	6.00	6.00	6.00	6.00	4.00	6.00
26	Death	A very young patient dies	5	3	0.00	4.83	6.67	5.00	2.67	5.00	3.67	6.00
27	Death	An unusual number of patients die in a short period of time	6	3	0.59	4.33	6.00	4.67	2.67	3.33	3.67	5.67
28	Death	You provide treatment to a patient who is about to die	5	3	0.00	3.89	5.33	4.67	1.33	4.00	2.67	5.33
29	Death	A patient commits suicide during your shift	11	1	NA	6.50	7.00	7.00	6.00	6.00	6.00	7.00
30	Death	You immediately go back to work after experiencing a difficult event (such as	4	3	0.00	4.28	5.33	4.67	2.67	4.00	3.33	5.67

Table 8 - Continued

		the sudden death of a patient)										
31	Emotional labor	You have to hide your true feelings	3	3	0.00	5.33	6.00	6.00	4.00	5.33	4.67	6.00
32	Emotional labor	You are expected to be very "positive" with a patient, although you don't want to be	6	3	0.00	5.44	6.67	5.00	3.33	6.67	6.67	4.33
33	Emotional labor	You are unable to express an angry feeling	5	3	0.00	3.61	4.33	4.00	2.67	3.67	2.67	4.33
34	Emotional labor	You emotionally support a patient	5	4	0.00	3.83	5.00	3.75	2.50	3.75	4.00	4.00
35	Emotional labor	You calm down a patient	3	3	0.00	4.39	5.67	4.00	2.33	4.33	5.67	4.33
36	Family issues	Family expectations increase because a patient who is about die suddenly improves but you know it is only temporary	6	3	0.91	5.17	6.00	5.67	2.33	5.33	6.00	5.67
37	Family issues	A patient's family member physically attacks you	5	2	0.00	6.50	7.00	7.00	7.00	7.00	7.00	4.00
38	Family issues	A patient's family member physically attacked another staff member	7	3	0.00	4.78	4.33	4.00	4.67	6.67	4.67	4.33
39	Family issues	A patient's family member is disrespectful to you	4	4	0.00	4.63	4.00	5.00	4.00	5.75	5.00	4.00
40	Family issues	A patient's family member has multiple requests from you	3	3	0.00	3.39	3.33	2.67	3.00	4.00	4.33	3.00

Table 8 - Continued

41	Family issues	A patient's family member is constantly present, almost as if he/she works in your department	4	3	0.00	4.44	4.33	4.33	4.00	5.33	4.33	4.33
42	Family issues	A patient's family member is very negative and makes the patient feel worse	5	4	0.00	4.63	4.75	4.00	4.00	5.50	5.50	4.00
43	Family issues	A guardian denies provision of treatment you know will help due to religious beliefs or ideology	4	3	0.00	4.39	5.33	4.00	1.67	5.00	4.67	5.67
44	Family issues	You witness a family conflict	6	3	0.00	3.83	3.67	2.67	3.00	5.67	4.67	3.33
45	Family issues	A patient's family member enters your break/lunch room	4	3	0.77	5.33	5.00	6.33	5.00	5.67	5.00	5.00
46	Family issues	Family members are not supportive enough to your child patient	6	1	NA	5.17	7.00	2.00	2.00	7.00	6.00	7.00
47	Family issues	Family members are not supportive enough to your elderly patient	4	3	0.15	4.89	5.67	4.00	2.67	5.67	5.67	5.67
48	Family issues	Crowding is interfering with equipment mobilization	4	3	0.00	3.89	3.67	4.67	2.67	4.33	3.33	4.67
49	Family issues	You are in a closed room with a patient and family member	3	3	0.49	2.00	2.67	1.67	1.00	2.33	2.33	2.00

Table 8 - Continued

50	Family issues	You suspect that you child patient is being neglected by their parent	5	3	0.59	5.33	6.33	5.33	1.67	6.33	6.33	6.00
51	Family issues	Patient parents who are addicted to drugs arrived under the influence of drugs	4	3	0.00	5.17	5.67	4.00	4.33	7.00	5.00	5.00
52	Family issues	You suspect that your patient suffers from abuse or neglect at home	3	3	0.00	5.22	6.00	4.67	3.33	6.00	5.33	6.00
53	Family issues	A patient's family member tries to manage you	5	4	0.00	3.63	4.50	3.00	2.25	4.50	3.50	4.00
54	Family issues	A patient's family member is nagging you	3	3	0.00	5.11	5.67	5.33	3.67	6.00	5.00	5.00
55	Family issues	A patient's family member is talking to you while you are working on the computer	5	4	0.04	2.13	1.50	3.00	1.50	2.25	3.00	1.50
56	Family issues	A patient's family member is intervening in the treatment you are providing	4	3	0.00	5.00	5.33	4.33	4.00	5.33	5.67	5.33
57	Family issues	A patient's family member is yelling at you	4	4	0.93	6.42	6.25	6.50	6.75	6.75	5.50	6.75
58	Family issues	A patient's family member is threatening you	4	3	0.81	6.50	6.33	7.00	6.33	7.00	6.00	6.33
59	Family issues	A patient's family member asks you a lot of questions	4	3	0.00	4.72	5.00	5.33	3.00	5.67	6.00	3.33

Table 8 - Continued

60	Family issues	You asked the family members of your patient to leave the room	5	4	0.00	4.83	5.25	4.75	3.75	5.75	5.75	3.75
61	Legal	Your patient is audio recording you	7	3	0.00	3.56	3.67	3.33	3.67	4.00	3.00	3.67
62	Legal	Your patient implicitly threatens to sue you	6	3	0.84	6.17	6.33	6.33	6.00	6.67	5.67	6.00
63	Legal	You send someone for a test just because you are afraid of a lawsuit	9	4	0.83	5.38	5.00	5.25	5.25	5.75	5.75	5.25
64	Legal	Someone sues your employer because of your work	11	1	NA	4.83	5.00	3.00	5.00	5.00	6.00	5.00
65	Off-work events	You had to deal with a personal issue during work time	3	3	0.00	3.89	3.67	6.00	2.33	3.00	4.33	4.00
66	Off-work events	You got an urgent request to come to work	6	3	0.84	4.11	5.00	5.67	3.00	3.67	3.67	3.67
67	Patient condition	Your patient is a criminal	5	4	0.00	3.29	3.50	3.00	2.75	4.75	2.50	3.25
68	Patient condition	A patient moved to the front of the queue because of his/her medical condition	5	3	0.73	3.28	3.33	3.00	3.00	3.67	3.33	3.33
69	Patient condition	Your patient is not cooperating with the treatment	4	4	0.00	4.13	5.50	3.75	1.50	5.25	5.25	3.50
70	Patient condition	Your patient said he/she wants to die	6	4	0.00	4.04	4.50	3.75	2.00	3.50	5.75	4.75

Table 8 - Continued

71	Patient condition	A new patient just arrived and will likely need immediate CPR	6	4	0.66	3.38	5.50	4.25	2.00	2.25	1.25	5.00
72	Patient condition	Your young patient (child/baby) is withdrawing from alcohol/drugs	8	1	NA	2.67	2.00	3.00	3.00	3.00	2.00	3.00
73	Patient condition	Your patient is depressed	4	3	0.00	3.67	4.67	3.00	3.00	3.33	3.33	4.67
74	Patient condition	You provide CPR to the same patient again and again	7	3	0.00	3.72	5.33	4.00	1.67	4.00	2.33	5.00
75	Patient condition	Your patient has a medical condition because of caregiver neglect	4	3	0.59	4.44	5.33	4.33	2.00	5.00	4.00	6.00
76	Patient condition	You are treating a patient who is severely neglected (e.g. lice on an old woman / a baby)	5	3	0.00	4.11	5.00	3.67	2.00	4.67	4.33	5.00
77	Patient condition	Your patient is very frustrated with their medical condition	3	3	0.00	4.50	5.33	4.33	4.00	5.00	4.00	4.33
78	Patient condition	Your patient has a severe medical condition	3	3	0.00	3.67	5.00	3.67	2.67	3.33	2.67	4.67
79	Patient condition	Your patient has a complex medical condition	3	3	0.00	4.06	5.00	3.67	2.00	3.67	5.33	4.67
80	Patient condition	A patient you are treating is alone and has no support	5	4	0.45	4.13	5.25	4.25	2.25	3.75	4.00	5.25

Table 8 - Continued

81	Patient condition	You see a patient who suffers from severe pain, and you can't help	4	4	0.60	3.96	5.25	3.50	1.75	3.75	3.75	5.75
82	Patient condition	Your patient lost a lot of blood	7	3	0.00	2.44	4.00	1.67	1.00	4.00	1.00	3.00
83	Patient condition	Your patient's condition does not require medical attention	3	3	0.00	2.44	2.67	2.33	1.33	3.00	2.67	2.67
84	Patient condition	You suspect the medical diagnosis of a patient but cannot tell the patient until more test are run	4	4	0.71	3.92	4.50	3.75	2.50	4.00	4.00	4.75
85	Patient condition	You tie a patient to the bed to prevent self harm	6	3	0.97	6.22	7.00	6.33	5.00	6.00	6.33	6.67
86	Patient condition	You force hospitalization of a patient	6	3	0.00	3.94	4.00	3.33	3.00	5.00	4.00	4.33
87	Patient condition	Your patient keeps trying to get out of bed despite being instructed not to	6	3	0.00	5.00	4.67	4.67	3.33	6.67	6.33	4.33
88	Patient condition	Your patient is addicted to drugs	3	3	0.00	3.33	3.33	2.00	2.33	5.67	3.00	3.67
89	Patient condition	Your patient suffers from PTSD (Post Traumatic Stress Disorder)	6	3	0.00	4.78	6.00	4.00	4.33	6.00	4.67	3.67
90	Patient condition	Your patient does not arrive on time for treatment	4	4	0.00	2.13	1.25	3.00	1.00	3.25	2.50	1.75
91	Patient condition	You experience difficulties communicating with your patient	3	3	0.00	3.61	4.67	3.67	1.67	4.67	3.33	3.67

Table 8 - Continued

92	Patient expectations	Your patient is nagging you	3	3	0.00	3.17	3.33	2.33	2.00	4.67	4.67	2.00
93	Patient expectations	A patient refuses to be treated by you because of your gender	9	3	0.81	4.22	4.00	4.67	4.00	5.33	4.00	3.33
94	Patient expectations	A patient refuses to be treated by you because of your nationality	4	1	NA	4.83	4.00	4.00	5.00	7.00	5.00	4.00
95	Patient expectations	A patient refuses to be treated by you because of your race	10	3	0.00	3.72	2.67	4.00	3.67	6.00	2.33	3.67
96	Patient expectations	Your patient requests a different nurse or a doctor	9	3	0.00	3.94	3.33	3.67	4.00	4.67	4.00	4.00
97	Patient expectations	Your patient requests treatment elsewhere	4	3	0.00	4.72	4.67	4.00	4.00	6.33	6.00	3.33
98	Patient expectations	Several patients team up with complaints	5	3	0.00	4.11	4.33	3.67	4.33	4.33	3.67	4.33
99	Patient expectations	You are doing something just to meet a patient's expectation but it is irrelevant for treatment	5	4	0.00	3.00	3.00	2.75	1.75	4.75	3.75	2.00
100	Patient expectations	Your patient requests something immediately, and it is not possible	5	4	0.74	4.83	5.00	4.50	4.25	4.50	5.50	5.25
101	Patient expectations	Your patient expects an unrealistic outcome	3	3	0.00	4.94	4.67	5.00	4.00	5.67	6.33	4.00
102	Patient expectations	Your patient expects special treatment because of their social status	4	3	0.00	4.39	4.00	3.33	2.33	6.00	6.00	4.67

Table 8 - Continued

103	Patient expectations	A patient refuses treatment from you specifically	3	3	0.92	4.83	5.33	5.00	3.33	5.33	5.00	5.00
104	Patient expectations	You must deal with a dissatisfied patient	4	4	0.87	5.46	5.75	5.50	5.25	6.25	4.75	5.25
105	Patient expectations	Your patient accuses you of lying to him/her	4	3	0.00	5.22	5.00	4.67	5.00	6.00	5.67	5.00
106	Patient expectations	Your patient asks a lot of questions	4	3	0.00	3.33	4.00	3.33	2.33	4.33	3.67	2.33
107	Patient physical aggression	Your patient previously acted violently in your department	4	4	0.43	5.58	5.25	5.75	4.75	6.25	6.50	5.00
108	Patient physical aggression	A patient enters a treatment room where you are treating another patient	5	4	0.90	3.50	2.25	3.50	2.00	4.75	5.25	3.25
109	Patient physical aggression	A patient throws something at you	6	2	0.00	5.33	6.50	6.50	3.50	6.50	5.50	3.50
110	Patient physical aggression	A patient physically attacks another staff member	6	4	0.00	4.71	5.25	5.00	4.50	5.25	3.25	5.00
111	Patient physical aggression	A patient physically threatens you	4	4	0.89	6.29	6.50	6.00	6.50	6.25	6.00	6.50
112	Patient physical aggression	Security arrives because of a patient's aggressive behavior	7	3	0.89	6.22	6.33	5.67	6.00	7.00	6.33	6.00
113	Patient physical aggression	A patient intrudes into a private conversation you are having with another	5	4	0.23	4.21	3.50	5.25	2.50	6.25	3.75	4.00

Table 8 - Continued

		staff member and demands attention										
114	Patient physical aggression	A patient approaches your work station aggressively	4	4	0.97	6.63	6.50	6.75	6.75	7.00	6.50	6.25
115	Patient physical aggression	A patient threatens you with a weapon	10	1	NA	6.83	7.00	7.00	7.00	7.00	7.00	6.00
116	Patient verbal aggression	Your patient complains about you in front of other people	4	4	0.88	6.00	6.00	6.00	6.50	6.00	5.25	6.25
117	Patient verbal aggression	A patient disrespects you	3	3	0.00	4.83	5.00	6.00	4.67	4.33	4.00	5.00
118	Patient verbal aggression	A patient yells at you in front of other people	4	3	0.00	4.94	5.00	5.00	5.00	6.00	3.33	5.33
119	Patient verbal aggression	A patient yells at you	6	4	0.00	5.79	5.75	4.75	6.00	6.50	6.75	5.00
120	Patient verbal aggression	A patient yells in the department	4	3	0.00	4.22	3.67	3.67	3.00	5.33	5.67	4.00
121	Patient verbal aggression	A patient accuses you of not being professional	5	3	0.00	5.44	5.00	5.67	6.00	5.67	4.33	6.00
122	Patient verbal aggression	A patient curses you really badly	3	3	0.00	4.72	4.33	5.00	5.00	5.00	4.67	4.33

Table 8 - Continued

123	Personal staff member issues	You feel a sense of identification with a patient's family member	3	3	0.00	3.83	6.33	6.00	1.33	2.00	2.33	5.00
124	Personal staff member issues	You check your email repeatedly	4	4	0.95	1.33	1.50	2.00	1.00	1.50	1.00	1.00
125	Personal staff member issues	You have a sore throat and it makes it hard for you to talk	4	3	0.00	4.06	4.67	5.00	3.00	3.33	3.67	4.67
126	Personal staff member issues	You had to physically run from one place to another to complete tasks	4	4	0.00	2.75	3.00	3.75	2.25	3.00	2.75	1.75
127	Personal staff member issues	You feel a sense of personal identification with your patient (e.g., similar age, occupation)	3	2	0.00	1.92	1.50	3.00	1.00	3.50	1.50	1.00
128	Personal staff member issues	You are personally acquainted with a patient who arrives in your department	7	3	0.67	3.06	3.67	4.00	3.00	3.33	2.33	2.00
129	Personal staff member issues	You cry in front of your patient	8	1	NA	3.50	5.00	5.00	3.00	2.00	1.00	5.00
130	Personal staff	You arrive to your shift at the last minute and need to start to work immediately	3	3	0.55	2.00	1.67	3.00	1.67	2.00	1.67	2.00

Table 8 - Continued

	member issues											
131	Personal staff member issues	You feel like you are over-checking yourself	3	3	0.43	4.89	6.33	6.33	4.00	2.67	4.33	5.67
132	Professional challenge	You don't have a solution for your patient's needs	5	4	0.00	3.88	4.75	4.25	2.25	4.00	4.00	4.00
133	Professional challenge	A treatment you gave didn't work	3	3	0.00	2.56	3.00	3.00	2.67	2.00	2.00	2.67
134	Professional challenge	You encounter a medical condition you don't know enough about	3	3	0.00	4.17	5.33	6.33	2.67	2.67	3.33	4.67
135	Death	A CPR procedure you gave, failed	8	3	0.00	4.28	6.33	5.67	2.33	2.67	2.67	6.00
136	Professional challenge	You couldn't insert an IV properly	4	3	0.00	4.56	5.00	5.67	4.67	3.33	3.33	5.33
137	Professional challenge	You provided a treatment you are inexperienced in giving	5	3	0.90	4.17	5.33	4.33	3.33	3.00	4.67	4.33
138	Role conflict and overload	You are requested to do tasks that are not a part of your training (for example - making phone calls, cleaning the corridor)	4	1	NA	3.50	3.00	5.00	3.00	3.00	3.00	4.00
139	Professional challenge	You encounter an unusual medical condition that requires a lot of time from you	3	3	0.00	3.83	4.33	4.00	3.00	3.00	4.00	4.67

Table 8 - Continued

140	Professional challenge	You cannot take your eyes of a monitor you need to follow closely	6	3	0.00	3.44	4.00	3.33	2.67	3.67	4.00	3.00
141	Professional challenge	Your patient said that you don't know what you're talking about based on a Google search	4	3	0.00	4.89	3.67	4.67	5.00	6.67	5.67	3.67
142	Professional challenge	There is a task that you can't complete on time	4	3	0.00	3.33	3.00	4.33	2.00	3.67	4.00	3.00
143	Professional challenge	You told a staff member that they are not good at their job	3	3	0.55	3.83	6.00	5.33	1.33	2.33	2.00	6.00
144	Professional challenge	You think of your responsibility to “do no harm”	8	4	0.55	2.25	3.50	3.00	1.75	1.75	1.75	1.75
145	Professional challenge	Your patient needs a medical procedure that you are able, but not allowed, to perform	3	3	0.00	3.78	4.33	3.33	2.00	4.00	4.67	4.33
146	Professional challenge	You provide CPR to a patient	4	3	0.00	4.89	6.67	5.00	3.00	4.33	4.33	6.00
147	Professional challenge	There is a task that requires your immediate attention	3	3	0.00	3.89	4.33	3.00	3.00	5.33	4.33	3.33
148	Professional challenge	You must perform a task without clear guidelines	2	2	0.00	4.17	4.00	4.50	2.00	4.50	5.50	4.50
149	Professional challenge	You need to reprioritize your tasks	4	3	0.83	4.33	5.33	5.33	3.33	4.00	4.67	3.33

Table 8 - Continued

150	Professional challenge	You had to deliver bad news over the phone because of COVID-19	6	4	0.68	4.42	6.50	5.50	2.00	3.25	2.75	6.50
151	Professional challenge	You are not sure what medical procedure your patient needs	5	3	0.00	3.78	5.00	4.00	3.33	3.67	3.33	3.33
152	Professional challenge	A person collapsed in front of you, and you are not sure what to do	9	4	0.00	4.42	5.50	6.00	3.50	2.75	2.50	6.25
153	Professional challenge	A patient with COVID-19 is deteriorating in front of you, but she is in quarantine and you cannot support her	8	3	0.00	4.94	5.67	5.67	2.67	4.67	5.33	5.67
154	Professional challenge	Your performance is being evaluated by someone who is watching you	3	3	0.00	3.11	3.33	4.00	3.67	2.00	2.33	3.33
155	Co-worker issues	You experience a conflict with a staff member about a treatment	5	3	0.49	3.89	3.67	4.00	2.33	4.67	5.67	3.00
156	Co-worker issues	You and another staff member are having a private “venting” conversation	4	3	0.86	2.72	3.67	3.33	1.33	2.33	2.33	3.33
157	Co-worker issues	Another staff member shouted in your department	3	2	0.00	4.17	3.50	3.00	4.50	5.00	5.00	4.00
158	Co-worker issues	A staff member undermines you in front of other staff members	5	4	0.00	5.25	5.25	5.50	5.75	5.00	4.50	5.50

Table 8 - Continued

159	Co-worker issues	A staff member physically attacks another staff member	5	1	NA	5.67	6.00	5.00	4.00	7.00	5.00	7.00
160	Professional challenge	You have to perform bureaucratic work	4	3	0.84	1.50	1.00	2.67	1.00	1.67	1.33	1.33
161	Professional challenge	You had to manipulate a report to meet official requirments	5	3	0.00	5.44	6.67	5.00	4.67	5.00	5.33	6.00
162	Role conflict and overload	A staff member is calling your personal phone during work and you must answer it	5	3	0.91	3.89	4.67	4.67	3.00	4.00	3.67	3.33
163	Supervision issues	Your manager is annoyed that you did not notice another person's medical mistake	9	2	0.71	4.42	4.00	5.00	5.00	4.00	4.00	4.50
164	System issues	You feel like you are being blamed for problems you cannot solve	4	3	0.87	6.39	6.67	6.67	6.33	7.00	5.33	6.33
165	System issues	You feel like too many guidelines are restricting you	3	3	0.00	2.06	2.67	2.00	1.00	2.33	2.00	2.33
166	Co-worker issues	A staff member doesn't arrive at all to the shift where you were supposed to work together	5	3	0.77	5.06	5.33	5.33	5.00	5.33	4.67	4.67
167	Co-worker issues	Another staff member refuses to cooperate with you	3	2	0.35	5.67	6.00	6.50	5.50	5.00	6.00	5.00

Table 8 - Continued

168	Co-worker issues	A staff member physically attacks you	5	2	0.95	6.17	6.00	6.00	7.00	6.00	5.50	6.50
169	Co-worker issues	A staff member makes a mistake	4	4	0.00	3.46	3.25	3.75	3.25	3.75	3.50	3.25
170	Co-worker issues	Other staff members push you to complete a task quickly	3	3	0.00	3.33	4.00	3.33	2.00	4.00	3.67	3.00
171	Death	You provide treatment that will likely harm a patient who will die soon anyway	7	3	0.83	3.61	5.00	3.33	1.67	4.67	4.33	2.67
172	Death	You provide treatment that will not help a patient who will die soon anyway	4	4	0.00	3.63	4.75	4.50	2.25	3.00	2.50	4.75
173	Family issues	Family members are not supportive enough to your infant patient	3	3	0.43	4.94	4.67	5.33	2.67	5.00	5.67	6.33
174	Family issues	A patient's family member is upset about something you did	4	4	0.00	5.58	5.75	5.25	5.75	5.50	6.00	5.25
175	Family issues	You suspect that your child patient is being abused by a parent	6	3	0.87	5.67	6.33	6.00	2.67	6.33	6.33	6.33
176	Family issues	You have to deliver bad news to family members of your patient	5	3	0.00	4.56	6.33	5.00	1.33	4.00	5.00	5.67
177	Family issues	A patient's family member is complaining about you in front of others	5	3	0.00	5.06	5.00	4.67	5.33	5.33	5.33	4.67

Table 8 - Continued

178	Family issues	A patient's family member is complaining in front of you	3	3	0.00	4.22	4.67	4.33	4.00	4.00	4.67	3.67
179	Family issues	A patient's family member says that you are not professional	3	2	0.00	5.50	5.50	6.50	6.00	6.50	3.50	5.00
180	Legal	Your patient is video recording you	7	2	0.89	6.33	6.00	6.00	6.50	6.50	7.00	6.00
181	Legal	You think of your legal liability	6	4	0.00	2.83	3.25	3.25	2.50	2.50	2.75	2.75
182	Off-work events	A patient called your private phone and yelled at you	8	1	NA	5.00	5.00	4.00	3.00	6.00	6.00	6.00
183	Patient condition	Your patient “gave up” on himself/herself	4	3	0.35	3.72	5.00	5.00	2.33	3.33	2.67	4.00
184	Patient condition	Your patient has an allergic reaction	5	4	0.00	2.25	2.75	2.50	1.25	3.25	1.75	2.00
185	Patient condition	Your patient's medical condition is deteriorating	3	3	0.00	3.89	5.00	4.67	1.67	4.00	3.33	4.67
186	Patient condition	Your patient survived but their quality of life will be severely damaged	7	4	0.00	4.33	6.00	3.50	3.00	4.50	3.50	5.50
187	Patient condition	A patient with COVID-19 asks you to bring her water because she is alone in quarantine	9	3	0.96	1.28	1.33	1.33	1.00	1.33	1.00	1.67
188	Patient condition	Your patient removes his/her IV	4	3	0.00	5.06	5.00	4.33	4.33	6.00	5.67	5.00

Table 8 - Continued

189	Patient physical aggression	A person with a high social status is acting violently	6	3	0.00	5.22	5.33	4.33	5.00	6.33	6.00	4.33
190	Patient physical aggression	A patient physically attacks you	7	3	1.00	6.89	7.00	7.00	7.00	7.00	6.33	7.00
191	Professional challenge	You prepare a patient for surgery	4	3	0.00	2.22	2.00	1.67	1.00	3.33	3.00	2.33
192	Professional challenge	You are wondering whether an action you took is correct	3	3	0.00	5.78	6.67	7.00	6.33	3.67	4.33	6.67
193	Professional challenge	It is night time, and you called the Doctor on Duty	4	4	0.00	2.92	2.00	3.00	2.25	3.75	3.00	3.50
194	Professional challenge	You were assigned a new patient	4	3	0.84	2.39	3.33	2.00	1.33	2.00	4.33	1.33
195	Professional challenge	Your patient needs multiple professionals to treat him/her	4	3	0.00	3.78	3.00	4.00	2.67	4.67	5.00	3.33
196	Professional challenge	A CPR procedure you give takes a lot of time	9	2	0.00	5.42	6.50	6.00	4.50	5.50	4.00	6.00
197	Professional challenge	You deliver bad news to a patient	3	3	0.84	4.22	6.00	5.00	1.33	3.67	3.67	5.67
198	Professional challenge	You send your patient for tests only because you are afraid of legal consequences	9	3	0.00	2.11	2.00	2.00	2.00	2.33	2.67	1.67
199	Professional challenge	You provide treatment to multiple patients with complex medical conditions	4	3	0.00	3.39	5.33	2.67	2.00	4.67	2.67	3.00

Table 8 - Continued

200	Professional challenge	Someone knocks on your door while you are with a patient	3	3	0.75	1.67	2.00	1.33	1.00	2.33	2.33	1.00
201	Professional challenge	Multiple patients need immediate attention, but there is not enough staff so you have to prioritize	3	3	0.35	5.17	6.33	4.33	4.33	5.67	5.67	4.67
202	Professional challenge	You had to act in the “gray area” of official guidelines so that you could complete your tasks on time	4	3	0.00	4.39	5.33	5.00	4.33	3.00	4.00	4.67
203	Professional challenge	You give treatment to a patient with COVID-19	4	3	0.00	3.56	4.67	3.67	2.33	3.67	3.33	3.67
204	Professional challenge	You feel uncertain regarding what will happen in your shift	3	3	0.00	3.83	5.00	4.33	2.67	2.67	3.67	4.67
205	Professional challenge	You are assigned multiple new patients at the same time	3	3	0.87	3.72	4.67	4.33	3.33	2.67	3.33	4.00
206	Professional error	You almost gave your patient the wrong medicine accidently	4	3	0.00	4.83	5.33	6.00	4.33	2.67	4.00	6.67
207	Professional error	You gave your patient the wrong medicine accidently	11	2	0.86	4.17	5.50	6.00	3.50	2.50	1.50	6.00
208	Professional error	Someone else almost gave a patient the wrong medicine accidently	5	3	0.00	3.39	4.33	3.00	1.33	4.33	3.67	3.67
209	Professional error	Someone else gave a patient the wrong medicine accidently	5	3	0.00	4.06	3.67	3.67	3.67	4.67	4.33	4.33

Table 8 - Continued

210	Professional error	You made a mistake in diagnosing a medical condition	8	3	0.00	4.67	6.00	5.33	3.67	3.33	3.67	6.00
211	Professional error	You made a mistake in identifying your patient	8	1	NA	1.50	1.00	4.00	1.00	1.00	1.00	1.00
212	Professional error	You gave a treatment that made your patient's condition deteriorate	11	3	0.00	5.39	6.33	6.33	5.00	4.00	4.33	6.33
213	Professional error	You made a mistake	2	2	0.00	4.92	6.50	5.50	5.50	4.50	1.50	6.00
214	Role conflict and overload	You have multiple managers	7	3	0.92	3.33	3.00	4.33	3.33	3.00	3.00	3.33
215	Role conflict and overload	You don't have enough time to complete a task	3	3	0.00	3.61	3.33	4.33	3.67	3.00	3.00	4.33
216	Role conflict and overload	You are managing multiple processes at the same time	4	4	0.00	4.04	4.75	4.25	3.00	4.75	4.50	3.00
217	Role conflict and overload	You have too many tasks	3	3	0.63	3.17	5.67	4.00	2.33	2.00	1.33	3.67
218	Role conflict and overload	Multiple patients arrive to your department at the same time	3	3	0.59	4.56	5.33	5.00	3.33	4.67	4.33	4.67
219	Role conflict and overload	You write a letter for a patient	5	3	0.59	3.44	4.33	3.67	1.67	3.67	3.33	4.00

Table 8 - Continued

220	Role conflict and overload	Your shift has ended but you must continue working	3	3	0.88	3.78	4.33	6.00	3.00	2.67	2.33	4.33
221	Role conflict and overload	You call someone regarding your patient and he/she don't answer	3	3	0.00	3.11	2.00	5.00	2.33	3.33	3.33	2.67
222	Role conflict and overload	You finish your shift, and some of your own tasks are left for the next shift	6	4	0.00	2.42	2.75	3.00	1.75	2.25	2.25	2.50
223	Role conflict and overload	You are waiting to consult with a senior doctor	7	4	0.75	2.50	2.25	2.25	1.75	3.50	3.00	2.25
224	Supervision issues	You explain your actions to a manager	3	3	0.00	4.28	4.33	4.33	3.33	4.00	5.33	4.33
225	Supervision issues	Your manager mistreats you	8	4	0.00	5.33	5.75	5.00	5.50	5.50	4.50	5.75
226	Supervision issues	You request something from your manager and the request is rejected without explanation	5	3	0.00	4.83	5.33	6.00	4.00	4.00	5.00	4.67
227	Supervision issues	You warn your supervisors about a problem but they ignore you	3	3	0.00	3.33	3.67	3.00	2.00	4.00	4.67	2.67
228	Supervision issues	Your manager acts disrespectfully towards you	3	3	0.00	5.44	6.00	6.00	5.33	5.33	4.33	5.67
229	Supervision issues	Your manager yells at you	6	2	0.94	6.58	6.50	6.50	6.00	7.00	6.50	7.00
230	Supervision issues	Your manager tries to change your working	4	3	0.86	6.33	7.00	7.00	6.33	5.33	5.33	7.00

Table 8 - Continued

		conditions without your consent											
231	Supervision issues	Your manager is unfair to you	4	3	0.73	4.72	4.67	4.67	4.00	5.33	4.67	5.00
232	Supervision issues	Your manager criticizes you	3	3	0.00	5.50	6.00	6.00	5.00	5.67	4.33	6.00
233	Supervision issues	Your manager humiliates you in front of other people	9	1	NA	6.00	7.00	7.00	7.00	7.00	1.00	7.00
234	Supervision issues	Your manager does not back you up	3	3	0.83	5.72	6.00	5.67	6.00	5.67	5.00	6.00
235		You were blamed for another persons mistake	5	0	NA	6.47	6.80	6.80	6.20	6.00	6.40	6.60
236	System issues	A medical device that you need to use is not properly maintained	3	3	0.67	5.22	5.67	6.00	4.33	5.00	5.00	5.33
237	System issues	You cannot complete a task due to a lack of adequate equipment	4	3	0.00	3.06	2.67	5.67	1.00	3.33	4.00	1.67
238	System issues	Your shift is not properly prepared (for example - a medication is out of stock)	3	3	0.00	4.61	4.33	5.67	3.67	4.67	5.33	4.00
239	System issues	Another department receives higher scores for patient satisfaction	3	3	0.90	2.39	2.00	2.33	2.33	2.67	2.33	2.67
240	System issues	You are the “face” of a treatment delay for which you are not responsible (e.g. a late ambulance or a doctor who hasn't arrived)	3	3	0.00	5.06	5.00	5.33	4.00	5.33	5.33	5.33

Table 8 - Continued

241	System issues	There is a recurring negative event that could have been avoided	6	3	0.00	5.56	6.67	5.33	3.67	6.00	4.67	7.00
242	System issues	Your shift is under staffed	3	3	0.00	3.39	4.33	5.00	2.00	2.00	3.33	3.67
243	System issues	There are patients "chattering" outside the treatment room you are working in	3	3	0.89	2.83	2.67	3.67	2.33	2.33	3.33	2.67
244	System issues	There is a long waiting time in your department	3	3	0.00	2.83	2.67	2.33	2.00	4.67	3.33	2.00
245	System issues	You are splitting your attention between multiple departments which makes you feel like you are doing a poor job in all places	4	3	0.00	4.78	5.00	5.33	4.33	4.00	4.67	5.33
246	System issues	There are many new staff members in your shift	4	3	0.00	2.67	3.00	2.67	1.67	3.33	3.00	2.33
247	System issues	There was a change in guidelines	4	3	0.67	2.17	2.00	3.00	2.00	2.00	2.00	2.00
248	System issues	You have no ability to take a vacation on your own terms (e.g., number of days)	4	3	0.00	4.11	4.00	6.33	2.67	4.33	4.00	3.33
249	System issues	You are doing administrative work (for example, you have to make a lot of phone calls to coordinate treatment)	3	3	0.00	2.28	2.33	3.00	1.33	3.00	2.67	1.33

Table 8 - Continued

250	System issues	Your tasks are distributed between distant locations	6	4	0.88	1.79	2.00	2.75	1.50	1.25	1.50	1.75
251	System issues	Your department was reorganized	4	3	0.00	5.39	5.67	6.33	4.67	5.00	5.67	5.00
252	System issues	The cafeteria at your work isn't open today	7	2	0.00	2.17	2.50	2.50	2.50	1.50	2.50	1.50
253	System issues	You need to move a patient to a different treatment room, but the room in not available	4	3	0.00	5.11	4.00	6.67	3.00	6.33	5.67	5.00
254	System issues	You need to request something from a different ward	3	3	0.00	3.00	2.00	2.00	2.33	4.33	5.00	2.33
255	System issues	Your patient need to be transferred but there are no staff available to do this	3	3	0.00	4.72	6.33	4.00	1.67	4.67	6.67	5.00
256	System issues	You cannot get the professional assistance you need on time	3	3	0.00	4.33	4.33	5.33	2.33	4.33	4.67	5.00
257	System issues	Your patient should be in a different department	3	3	0.00	2.22	2.00	2.00	2.00	2.00	3.67	1.67
258	System issues	You provide treatment that will not help your patient, but you have to do it because of protocol	4	3	0.00	4.17	5.00	2.67	2.33	4.33	5.00	5.67
259	System issues	You have to work without proper equipment that should protect you from COVID-19	4	3	0.75	5.78	6.00	6.67	6.00	4.33	5.67	6.00

Table 8 - Continued

260	Death	Your patient dies as a result of unprofessional treatment provided by others	8	3	0.35	6.11	7.00	6.67	4.67	6.00	5.33	7.00

Appendix 2–Precision and Recall of Sentiment Analysis Tools (Yom-Tov et al. 2018)

Table 9 - Precision and Recall of Sentiment Analysis Tools (Yom-Tov et al. 2018). Comparing four models in detecting emotion in customer messages.

Emotion class	Model	Precision	Recall	F_1	$F_{0.5}$
Negative	SentiStrength	0.494	0.216	0.300	0.393
	CustSent	0.719	0.236	0.355	0.51
	Stanford	0.335	0.509	0.404	0.36
	LIWC	0.479	0.115	0.186	0.294
Positive	SentiStrength	0.813	0.677	0.739	0.781
	CustSent	0.866	0.569	0.687	0.784
	Stanford	0.546	0.339	0.418	0.486
	LIWC	0.491	0.717	0.583	0.524

Appendix 3–Correlation Table: Message Level

Table 10 - Pairwise Pearson Correlation: Message Level

		1	2	3	4	5	6	7	8	9	10	11
1	EMO $(SentiStrength)$	1.00										
2	$CustSent$	0.61‡	1.00									
3	RT	-0.04‡	-0.04‡	1.00								
4	$log(RT)$	-0.06‡	-0.07‡	0.84‡	1.00							
5	$NumWords$	0.07‡	0.09‡	0.46‡	0.40‡	1.00						
6	$log(NumWords)$	0.06‡	0.08‡	0.39‡	0.42‡	0.89‡	1.00					
7	$Concurrent$	0.00	0.01‡	-0.00†	0.03‡	-0.05‡	-0.03‡	1.00				
8	$NumInQueue$	0.01‡	0.01‡	0.01‡	0.01‡	0.01‡	0.01‡	0.14‡	1.00			
9	$ConvStage$	0.25‡	0.33‡	0.12‡	0.07‡	0.22‡	0.21‡	-0.05‡	0.02‡	1.00		
10	$ShiftTime$	0.00	0.00	-0.01‡	-0.01‡	-0.01‡	-0.01‡	-0.02‡	0.01‡	0.02‡	1.00	
11	$Turn$	0.09‡	0.12‡	-0.03‡	-0.05‡	0.01‡	0.01‡	-0.19‡	-0.02‡	0.42‡	0.05‡	1.00

Note: $\dagger p < 0.05$, $\ddagger p < 0.01$.

Appendix 4–Correlation Table: Conversation Level

Table 11 - Pairwise Pearson Correlation: Conversation Level

		1	2	3	4	5	6	7	8	9	10	11	12	13
1	$NTurns$	1.00												
2	$log(NTurns)$	0.89‡	1.00											
3	EMO_1^a	-0.02‡	-0.01‡	1.00										
4	$CustSent_1$	-0.06‡	-0.06‡	0.43‡	1.00									
5	RT	-0.05‡	-0.06‡	-0.02‡	-0.04‡	1.00								
6	$log(RT)$	0.05†	0.08‡	-0.02‡	-0.05‡	0.94‡	1.00							
7	$NumWords$	-0.11‡	-0.10‡	0.01‡	0.01†	0.29‡	0.28‡	1.00						
8	$log(NumWords)$	-0.05‡	-0.01‡	0.01†	0.00	0.27‡	0.29‡	0.96‡	1.00					
9	$Concurrent$	-0.11‡	-0.08‡	0.01‡	0.01‡	0.04‡	0.05‡	-0.02‡	-0.02‡	1.00				
10	$NumInQueue$	-0.03‡	-0.03‡	0.01‡	0.00	0.01‡	0.01‡	0.02‡	0.02‡	0.09‡	1.00			
11	$CustWords_1$	-0.02‡	-0.01‡	0.00	-0.08‡	0.10‡	0.12‡	0.13‡	0.13‡	0.06‡	0.05‡	1.00		
12	$log(CustWords_1)$	-0.06‡	-0.05‡	0.00	-0.07‡	0.10‡	0.13‡	0.14‡	0.14‡	0.06‡	0.04‡	0.82‡	1.00	
13	$ShiftTime$	-0.00	-0.00	-0.01‡	-0.01‡	-0.02‡	-0.02‡	-0.02‡	-0.02‡	0.00	0.01‡	-0.01†	-0.01†	1.00

Note: †$p < 0.05$, ‡$p < 0.01$, $^a SentiStrength$.

Appendix 5–Analyses using OLS, No Instrumental Variables

Table 12 - Effect of Customer Emotion on Agent Behavior (Outliers Excluded, OLS with no IVs)

	Model(1) $log(RT)$	Model(2) $log(RT)$	Model (3) $log(NumWords)$	Model (4) $NTurns$
EMO_{t-1}	-0.096***	-0.097***	0.001	
	(0.002)	(0.001)	(0.001)	
EMO_1				-0.250***
				(0.030)
$Concurrent_t$	0.056***	0.073***	-0.040***	-1.238***
	(0.003)	(0.002)	(0.002)	(0.037)
$NumInQueue_t$	0.003***	0.003***	0.002*	0.030***
	(0.001)	(0.001)	(0.001)	(0.005)
$ConvStage_t$	0.168***	-0.040***	0.467***	
	(0.004)	(0.004)	(0.004)	
$log(NumWords_t)$		0.446***		
		(0.001)		
$log(CustWords_1)$				-0.328***
				(0.021)
$IsWeekend$				-0.017
				(0.041)
$SrvType$				5.975
				(3.120)
$ShiftTime$				Included
$HourOfDay$				Included
Conversation Fixed Effect	Included	Included	Included	
Agent Fixed Effect				Included
Constant	3.634***	2.247***	3.112***	11.102***
	(0.007)	(0.008)	(0.006)	(1.661)
Observations	650,856	650,856	650,856	141,654

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix 6–Analyses using Alternative Approach to Reduce Measurement Error using Factor Analysis

Table 13 - Effect of Customer Emotion on Agent Behavior (Outliers Excluded, EMO_FA_{t-1} is based on Factor Analysis of SentiStrength and CustSent)

	Model(1) $log(RT)$	Model (3) $log(NumWords)$	Model (2) $log(RT)$
EMO_FA_{t-1}	-0.124***	0.003†	-0.125***
	(0.0016)	(0.0064)	(0.0015)
$Concurrent_t$	0.056***	-0.040***	0.074***
	(0.0026)	(0.0024)	(0.0024)
$NumInQueue_t$	0.003***	0.002*	0.003***
	(0.0007)	(0.0006)	(0.0006)
$ConvStage_t$	0.215***	0.465***	0.007***
	(0.0040)	(0.0037)	(0.0040)
$log(NumWords_t)$			0.446***
Conversation Fixed Effect	Included	Included	Included
Constant	3.570***	2.787***	1.753***
	(0.007)	(0.0064)	(0.0100)
Indirect Effects			
EMO_FA_{t-1} via $log(NumWords_t)$			0.0012†
			(0.0007)
Observations	650,856	650,856	650,856

Standard errors in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, †$p < 0.1$

Appendix 7–Analyses With and Without Log Transformation

Table 14 - Effect of Customer Emotion on Agent Behavior (Outliers Excluded, Both EMO_1 and EMO_{t-1} are Instrumented Using CustSent. Models (1)–(3) are Without log transformations of the DVs, Model (4) is with log transformation of the DV)

	Model(1) RT	Model(2) RT	Model (3) $NumWords$	Model(4) $log(Nturns)$
EMO_1				-0.049***
				(0.084)
EMO_{t-1}	-12.236***	-13.728***	1.454***	
	(0.221)	(0.201)	(0.091)	
$Concurrent_t$	3.125***	4.836***	-1.668***	
	(0.208)	(0.189)	(0.085)	
$Concurrent$ (chat level)				-0.102***
				(0.003)
$NumInQueue_t$	0.160**	0.122*	0.036	
	(0.056)	(0.051)	(0.023)	
$NumInQueue$ (chat level)				0.003***
				(0.000)
$ConvStage_t$	27.720***	11.087***	16.213***	
	(0.335)	(0.308)	(0.138)	
$log(NumWords_t)$		1.026***		
		(0.003)		
$log(CustWords_1)$				-0.024***
				(0.002)
$IsWeekend$				-0.005
				(0.003)
$SrvType$				0.531*
				(0.265)
$ShiftTime$				Included
$HourOfDay$				Included
Agent Fixed Effect				Included
Constant	44.666***	15.427***	28.502***	2.178***
	(0.552)	(0.510)	(0.227)	(0.141)
Observations	650,856	650,856	650,856	141,654

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix 8–Analyses using Agent Emotion as an Additional Mediator

Table 15 - Effect of Customer Emotion on Agent Behavior (Outliers Excluded, EMO_{t-1} is Instrumented using $CustSent_{t-1}$)

	Model (3) $log(NumWords)$	Model (3)[1] $AgentEmo$	Model (2) $log(RT)$	Model (2) $log(RT)$
EMO_{t-1}	0.007**	0.210***	-0.155**	-0.197***
	(0.002)	(0.002)	(0.002)	(0.002)
$log(NumWords_t)$			0.468***	0.446***
			(0.001)	(0.001)
$AgentEmo_t$			-0.199***	
			(0.002)	
$Concurrent_t$	-0.040***	0.002	0.076***	0.074***
	(0.002)	(0.002)	(0.002)	(0.002)
$NumInQueue_t$	0.002*	0.001	0.003***	0.003***
	(0.001)	(0.001)	(0.001)	(0.001)
$ConvStage_t$	0.464***	0.280***	0.055***	0.006†
	(0.004)	(0.003)	(0.004)	(0.004)
Conversation Fixed Effect	Included	Included	Included	Included
Constant	2.764***	0.209***	1.764***	1.809***
	(0.006)	(0.005)	(0.001)	(0.001)
Indirect Effects				
EMO_{t-1} via $log(NumWords_t)$			0.003**	0.003**
			(0.001)	(0.001)
EMO_{t-1} via $AgentEmo_t$			-0.042***	
			(0.001)	
Observations	650,856	650,856	650,856	650,856

Note.[1]In the second column, we used Model 3 with a different DV ($AgentEmo_t$), and in the third column we used Model 2 including $AgentEmo_t$.
Standard errors in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, †$p < 0.1$

Appendix 9–Analyses using Alternative Measures of Concurrency

Table 16 - Effect of Customer Emotion on Agent Behavior (Outliers Excluded, EMO_{t-1} is Instrumented using $CustSent_{t-1}$)

	Model(1) $log(RT)$	Model (3) $log(NumWords)$	Model (2) $log(RT)$	Model(1) $log(RT)$	Model (3) $log(NumWords)$	Model (2) $log(RT)$
EMO_{t-1}	-0.204***	0.007**	-0.194***	-0.205***	0.007**	-0.195***
	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)
$Concurrent\_words_t$	0.009***	0.002***	0.008***			
	0	0	0			
$Concurrent\_msg_t$				0.419***	0.095***	0.382***
				(0.001)	(0.001)	(0.001)
$NumInQueue_t$	0	0	0.001***	0	0	0
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$ConvStage_t$	0.184***	0.460***	-0.033***	0.176***	0.458***	-0.035***
	(0.004)	(0.004)	(0.003)	(0.004)	(0.004)	(0.003)
$log(NumWords_t)$			0.402***			0.391***
			(0.001)			(0.001)
Conversation Fixed Effect	Included	Included	Included	Included	Included	Included
Constant	3.616***	2.719***	2.184***	3.571***	2.718***	2.16***
	(0.003)	(0.005)	(0.009)	(0.003)	(0.005)	(0.008)
Indirect Effects						
EMO_{t-1} via $log(NumWords_t)$			0.003***		0.003***	
			(0.001)		(0.001)	
Observations	651,709	651,709	651,709	651,709	651,709	651,709

Standard errors in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix 10–Analyses using Clustered Standard Errors

Table 17 - Effect of Customer Emotion on Agent Behavior (Outliers Excluded, EMO_{t-1} is Instrumented using $CustSent_{t-1}$. First column is with clustered standard errors. Second column is the original model.)

	Model (1) $log(RT)$	Model (1) $log(RT)$
EMO_{t-1}	-0.206***	-0.206***
	(0.003)	(0.003)
$Concurrent_t$	0.057***	0.057***
	(0.003)	(0.003)
$NumInQueue_t$	0.003***	0.003***
	(0.001)	(0.001)
$ConvStage_t$	0.246***	0.246***
	(0.005)	(0.005)
Conversation Fixed Effect	Included	Included
Constant	3.617***	3.617***
	(0.007)	(0.007)
Observations	650,856	650,856

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix 11–Survival Analysis with Categorical Emotion

Table 18 - Effect of Customer Emotion on the Length of a Conversation (Outliers Excluded, In the first, fourth, and fifth columns. EMO Variables are Instrumented using CustSent.)

	Model (4) $Nturn$	Model (5) $Pr(LastTurn)$	Model (5) $Pr(LastTurn)$	Model (5) $Pr(LastTurn)$	Model (5) $Pr(LastTurn)$
$EMO_1\_positive$	-0.726***		-0.133***		-0.265***
	(0.1232)		(0.0049)		(0.0118)
$EMO_1\_negative$	3.463***		-0.069***		-0.339***
	(0.1766)		(0.0062)		(0.0154)
$EMO\_positive_{t-1}$		0.443***	0.478***	0.872***	0.920***
		(0.0043)	(0.0044)	(0.0063)	(0.0065)
$EMO\_negative_{t-1}$		0.097***	0.118***	0.061**	0.166***
		(0.0067)	(0.007)	(0.0196)	(0.0204)
$Concurrent_t$		0.021***	0.021***	0.018***	0.019***
		(0.0031)	(0.0031)	(0.0031)	(0.0031)
$NumInQueue_t$		0.001	0.001*	0.001	0.001*
		(0.0005)	(0.0005)	(0.0005)	(0.0005)
$log(CustWords_t)$		-0.060***	-0.057***	-0.062***	-0.058***
		(0.0019)	(0.0019)	(0.0021)	(0.0021)
$log(CustWords_1)$	-0.420***				
	(0.0239)				
$NumInQueue$ (chat level)	0.030***				
	(0.0047)				
$Concurrent$ (chat level)	-1.238***				
	(0.0369)				
$Turn_t$		-0.002***	-0.002***	-0.007***	-0.006***
		(0.0005)	(0.0005)	(0.0005)	(0.0006)
$IsWeekend$	-0.023	0.009*	0.009	0.009*	0.010*
	(0.0416)	(0.0046)	(0.0046)	(0.0045)	(0.0045)
$SrvType$	6.473*	-0.306*	-0.329*	-0.286*	-0.353**
	(3.1506)	(0.1308)	(0.1306)	(0.1295)	(0.1298)
$ShiftTime$	Included	Included	Included	Included	Included
$HourOfDay$	Included	Included	Included	Included	Included
Agent Fixed Effect	Included	Included	Included	Included	Included
Constant	11.606***	-0.857***	-0.833***	-1.225***	-1.147***
	(1.6853)	(0.1463)	(0.1464)	(0.1773)	(0.1773)
Observations	141,654	518,437	518,437	518,437	518,437

Standard errors in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix 12–Analyses Including Outliers in the Sample (Continuous *EMO*)

Table 19 - Effect of Customer Emotion on Agent Behavior (Outliers Included, Both EMO_1 and EMO_{t-1} are Instrumented using CustSent)

	Model(1) $log(RT)$	Model(2) $log(RT)$	Model (3) $log(NumWords)$	Model (4) $NTurns$
EMO_{t-1}	-0.343***	-0.332***	-0.024***	
	(0.0031)	(0.0030)	(0.0023)	
EMO_1				-1.677***
				(0.0638)
$Concurrent_t$	0.098***	0.112***	-0.031***	-1.153***
	(0.0030)	(0.0028)	(0.0022)	(0.0337)
$NumInQueue_t$	0.004***	0.003***	0.002***	0.030***
	(0.0008)	(0.0008)	(0.0006)	(0.0043)
$ConvStage_t$	0.280***	0.016***	0.570***	
	(0.0048)	(0.0046)	(0.0035)	
$log(NumWords_t)$		0.462***		
		(0.0016)		
$log(CustWords_1)$				-0.299***
				(0.0196)
$IsWeekend$				0.0171
				(0.0385)
$SrvType$				5.612
				(3.1187)
$ShiftTime$				Included
$HourOfDay$				Included
Conversation Fixed Effect	Included	Included	Included	
Agent Fixed Effect				Included
Constant	3.442***	2.018***	3.083***	10.716***
	(0.0079)	(0.0089)	(0.0058)	(1.5582)
Observations	825577	825577	825583	162362

Standard errors in parentheses * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix 13–Analyses Including Outliers in the Sample (Categorical EMO)

Table 20 - Effect of Customer Emotion on Agent Behavior (Outliers Included, All EMO Variables are Instrumented usingCustSent)

	Model(1) $log(RT)$	Model(2) $log(RT)$	Model (3) $log(NumWords)$	Model (4) $NTurns$
$EMO\_positive_{t-1}$	-0.525***	-0.512***	-0.028***	
	(0.0046)	(0.0043)	(0.0034)	
$EMO\_negative_{t-1}$	0.183***	0.121***	0.133***	
	(0.0143)	(0.0134)	(0.0105)	
$EMO\_positive_1$				-0.907***
				(0.1140)
$EMO\_negative_1$				3.271***
				(0.1638)
$Concurrent_t$	0.100***	0.114***	-0.031***	-1.154***
	(0.0030)	(0.0028)	(0.0022)	(0.0337)
$NumInQueue_t$	0.004***	0.003***	0.002***	0.029***
	(0.0008)	(0.0008)	(0.0006)	(0.0044)
$ConvStage_t$	0.315***	0.050***	0.574***	
	(0.0049)	(0.0047)	(0.0036)	
$log(NumWords_t)$		0.463***		
		(0.0016)		
$log(CustWords_1)$				-0.371***
				(0.0222)
$IsWeekend$				0.007
				(0.0386)
$SrvType$				5.868
				(3.1237)
$ShiftTime$				Included
$HourOfDay$				Included
Conversation Fixed Effect	Included	Included	Included	
Agent Fixed Effect				Included
Constant	3.481***	2.058***	3.075***	9.734
	(0.0080)	(0.0090)	(0.0059)	(7.0542)
Observations	825,577	825,577	825,583	162,362

Standard errors in parentheses * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix 14–Analyses Including Outliers in the Sample (Models 6 and 7)

Table 21 - Effect of Agent Behavior on Customer Emotion (Outliers Included. $log(RT_{t-1})$ is Instrumented using $Concurrent_{t-1}$ and $NumInQueue_{t-1}$)

	Model (6) EMO	Model (7) EMO
$log(RT_{t-1})$	-0.056***	-0.328***
	(0.0100)	(0.0256)
$ConvStage_t$	0.875***	1.114***
	(0.0044)	(0.0068)
$log(NumWords_{t-1})$		0.163***
		(0.0125)
$Turn_t$		-0.015***
		(0.0003)
Conversation Fixed Effect	Included	Included
Constant	0.055	0.528***
	(0.0354)	(0.0558)
Observations	725,805	725,805

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix 15–Analyses using OLS, No Instrumental Variables (DV: *EMO*)

Table 22 - Effect of Agent Behavior on Customer Emotion (Outliers Excluded)

	Model (6) EMO	Model (7) EMO
$log(RT_{t-1})$	0.022***	0.022***
	(0.001)	(0.001)
$ConvStage_t$	0.846***	1.133***
	(0.003)	(0.006)
$log(NumWords_{t-1})$		-0.009***
		(0.001)
$Turn_t$		-0.015***
		(0.000)
Conversation Fixed Effect	Included	Included
Constant	-0.2198***	-0.222***
	(0.005)	(0.005)
Observations	776,551	776,551

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix 16–Analyses using OLS, No Instrumental Variables (DV: *CustSent*)

Table 23 - Effect of Agent Behavior on Customer Emotion (Outliers Excluded, using CustSent as the Main Measure of Customer Emotion)

	Model (6) $CustSent$	Model (7) $CustSent$
$log(RT_{t-1})$	0.039***	0.035***
	(0.001)	(0.001)
$ConvStage_t$	1.015***	1.366***
	(0.003)	(0.006)
$log(NumWords_{t-1})$		-0.004**
		(0.001)
$Turn_t$		-0.019***
		(0.000)
Conversation Fixed Effect	Included	Included
Constant	-0.404***	0.414***
	(0.005)	(0.005)
Observations	776,551	776,551

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix 17-Analyses using *CustSent* as the Main Emotion Measure

Table 24–Effect of Customer Emotion on Agent Behavior (Outliers Excluded, using CustSent as the Main Measure of Customer Emotion)

	Model(1) $log(RT)$	Model(1) $log(RT)$	Model (3) $log(NumWords)$	Model (4) $Nturn$
$CustSent_1$				-0.710***
				(0.084)
$CustSent_{t-1}$	-0.184***	-0.185***	0.002	
	(0.003)	(0.003)	(0.003)	
$Concurrent_t$	0.058***	0.075***	-0.040***	
	(0.003)	(0.002)	(0.002)	
$Concurrent$ (chat level)				-1.231***
				(0.037)
$NumInQueue_t$	0.003***	0.002***	0.002*	
	(0.001)	(0.001)	(0.001)	
$NumInQueue$ (chat level)				0.029***
				(0.005)
$ConvStage_t$	0.273***	0.064***	0.466***	
	(0.005)	(0.004)	(0.004)	
$log(NumWords_t)$		0.446***		
		(0.001)		
$log(CustWords_1)$				-0.356***
				(0.021)
$IsWeekend$				-0.017
				(0.041)
$SrvType$				5.937
				(3.114)
$ShiftTime$				Included
$HourOfDay$				Included
Agent Fixed Effect				Included
Constant	3.584***	2.194***	3.113***	11.138***
	(0.007)	(0.008)	(0.006)	(1.658)
Observations	650,856	650,856	650,856	141,654

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix 18–Arrival Rate

This section presents additional data needed for the calculations of the impact of the results on offered load. Table 25 provides a typical pattern of customer arrival rate per working hour in our data.

Table 25 – Arrival Rate during a Working Day

Hour	λ
9	62.45
10	67.6
11	63.75
12	67.25
13	68.9
14	69.05
15	86.25
16	90.2
17	68.1
18	63.5
19	59.35
20	65.5
21	57.95
22	53.65

המגבלה המרכזית של מחקרים 1 ו -2 היא שהם לא בחנו את השפעתו של עומס רגשי על ביצועי עובדים. לכן, מטרת מחקר 3 הייתה לבחון השפעה אפשרית זו. במחקר 3 נותחו 141,654 שיחות שירות לקוחות אשר התקיימו בכתב (״צ׳אט״, שִׂיחוּחַ) בין עובדים של חברת תעופה מערבית לבין לקוחותיה. העומס הרגשי של עובדי השירות נמדד על ידי קידוד אוטומטי של הרגש (Sentiment Analysis) אותו הביעו לקוחות במהלך שיחת השירות (אחד האירועים שנמצא במחקר 2 כגורם מוסכם לעומס רגשי). בנוסף, נמדדו זמני התגובה של העובדים להודעות הלקוחות, מספר המילים בהם השתמשו העובדים (כאינדיקציה למאמץ שהשקיעו) וכמות החִזְרוּרים (איטרציות, פעולות החוזרות על עצמן) בין הודעות הלקוח להודעות העובד על להשלמת השירות. תוצאות המחקר מראות כי עומס רגשי הנוצר כתוצאה מרגש שלילי אותו הביעו לקוחות האריך את זמן התגובה של העובדים ואת מספר ההודעות הנחוצות להשלמת השירות. בנוסף, נמצא כי ישנה השפעה הדדית בין העומס הרגשי לזמני תגובה של העובדים: זמני תגובה גבוהים מעלים את הרגש השלילי שלקוחות מביעים, ולהפך.

תוצאות שלושת המחקרים מלמדות כי עומס רגשי קיים במערכות שירות שונות והוא נוצר כתוצאה מחשיפה של עובדים לאירועים נקודתיים. ישנם אירועים שנחווים באופן דומה על ידי עובדים, וזיהוי מערכתי של אירועים אלו עשוי לשפוך אור על מידת העומס הרגשי הנופל על העובדים, וזאת מבלי להשתמש בשאלוני דיווח-עצמי. מדידה רציפה של עומס רגשי של עובדים עשויה לסייע בניהול מערכתי של מערכות שירות ועובדי שירות על ידי ניתוב של עומס רגשי בצורה הוגנת בין עובדים, וסימון עובדים ומחלקות אשר נחשפו לרמות גבוהות של עומס רגשי לצורכי הקצאת הזדמנויות הפוגה ומתן כלים להתמודדות.

העבודה כוללת 3 מחקרים. מחקר 1 הינו מחקר איכותני המבוסס על ראיונות עומק עם 24 רופאים ואחיות במערכת הבריאות הציבורית והפרטית בישראל. במחקר זה, זוהו 260 אירועים אשר עלולים ליצר תחושת עומס או עומס רגשי בקרב עובדי שירות רפואי. האירועים כללו סוגים שונים של תופעות, כולל תופעות שנחשבו בעבר כגורמות לעומס רגשי ותופעות אחרות שלא נחשבו בעבר כגורמות לעומס רגשי, כמו למשל, קבלת מספר מטופלים חדשים במקביל. אולם, לא ניתן להסיק באיזה מידה האירועים שזוהו במחקר זה מהווים דרישה עבור כלל עובדי הצוותים הרפואיים.

לכן, המטרה המרכזית של מחקר 2 הייתה לאמוד את מידת הדרישה הרגשית של כל אירוע בנפרד, ואת היקף ההסכמה בין עובדים שונים על מידת הדרישה הרגשית. לשם כך גויס מדגם נוסף של 124 עובדי רפואה דוברי אנגלית ממדינות מערביות. כל משתתף במדגם זה התבקש לדרג 10 אירועים אשר נדגמו רנדומלית מתוך רשימת 260 האירועים שנאספו במחקר 1. עבור כל אירוע, כל משתתף דיווח את מידת העומס הרגשי אשר האירוע מייצר עבורו באמצעות סקאלה מתוקפת של 6 שאלות שהותאמו למדידה של אירוע בודד (לדוגמה, "מהי מידת הדרישה הרגשית של אירוע זה?"). כמו כן, כל משתתף דיווח מהי התדירות בה מופיע אירוע כזה במסגרת עבודתם השוטפת (החל מ"לעולם לא" ועד "מספר פעמים ביום"). תוצאות מחקר 2 הראו הסכמה בין העובדים שהשיבו על מידת העומס הרגשי של 53 אירועים מתוך 260. רשימת האירועים לגביהם ישנה הסכמה כוללת אירועים שמייצרים עומס רגשי מגוון, מנמוך עד גבוה, ושמתרחשים בתדירויות שונות. תוצאות המחקר השני מלמדות כי מגוון רחב של אירועים מייצר עומס רגשי לעובדי מערכות בריאות, לרבות אירועים אשר נחשבו בעבר כגורמים לעומס תפעולי בלבד (למשל, קבלת מספר מטופלים חדשים במקביל). כלומר מצאנו שתופעות אשר נחקרו בעבר בנפרד ולא סווגו כגורמות לעומס רגשי בכל זאת גורמות לעומס רגשי. ממצאי המחקר מלמדים הן על ההיקף הרחב של תופעות הגורמות לעומס רגשי, והן על מידת הדרישה הרגשית של אירועים שונים. בנוסף, מידת ההסכמה הגבוהה על מידת הדרישה הרגשית של 53 אירועים מלמדת כי עומס רגשי כולל רכיב אובייקטיבי, המשותף לעובדים רבים, בניגוד למחקרי עבר אשר מדדו תחושת סובייקטיביות של עומס רגשי בלבד. תוצאות מחקר 2 תומכות ברעיון כי ניתן למדוד את העומס הרגשי של עובדים באופן אשר מתקרב לאובייקטיביות על ידי קידוד האירועים אליהם נחשף העובד, וזאת במקום השימוש הנפוץ בשאלוני דיווח-עצמי.

תקציר

עומס במערכות שירות נחקר רבות תחת ההנחה שלקוחות הם הומוגניים ושעובדים במערכת פועלים באופן דומה אחד לשני. הנחה זו מופרת מפני ששני לקוחות המגיעים לקבל שירות זהה נבחנים אחד מהשני בבקשותיהם, ציפיותיהם, והאופן בו הם מתנהגים או מגיבים למצבים שונים. אותם לקוחות עשויים לגרום לרמות שונות של עומס על המערכת גם אם הם פנו לקבל שירות מאותה הסיבה. אחד מהגורמים המשפיעים על עבודתם של עובדי שירות הוא העומס הרגשי עמו עובדים מתמודדים. עומס רגשי עשוי לנבוע משלל גורמים בסביבתם של העובדים, למשל: סוג העבודה אותה הם מבצעים, מערכות יחסים של העובדים עם עמיתיהם ומנהליהם, התנהגותם של לקוחות וכן הלאה. גורמים אלו באים לידי ביטוי באירועים נקודתיים המתרחשים מעת לעת והעובדים חשופים אליהם (למשל, כאשר לקוח מרים את קולו בתהליך השירות).

הספרות המדעית אינה חד משמעית לגבי ההגדרה או לגבי אופן המדידה של עומס רגשי. ישנם מחקרים המניחים כי עומס רגשי נובע מדרישות הארגון מעובדיו להביע רגשות מסוימים (לדוגמא, רגש חיובי) ולהדחיק רגשות אחרים (לדוגמא, כעס). דרישה זו הוגדרה ונחקרה בעבר כ"עבודת רגשות", המוגדרת כחלק מתפקידם של עובדי קירות. מחקרים אחרים מתמקדים בסוגים אחרים של דרישות מעובדים, כמו למשל הדרישה מצוותי רפואה להתמודד עם דרישות בני משפחתם של מטופלים, ולספק טיפול למטופלים עם מחלה סופנית. ניכר כי המכנה המשותף בין מחקרים שונים על עומס רגשי הוא שיש דרישות מסוימות המייצרות עומס רגשי, אך לא ברור מהן אותן דרישות ואיך ניתן למדוד אותן.

בנוסף, מחקרי עבר מניחים כי עומס רגשי הינו חוויה סובייקטיבית והמדידה שלו נעשתה עד כה רק באמצעות דיווח-עצמי על שאלות כגון "מהי מידת הדרישה הרגשית של עבודתך?" מדידה זו אומדת חוויה סובייקטיבית של עומס רגשי במכלול היבטי העבודה, ולא מאפשרת זיהוי מצבים או אירועים נקודתיים שגורמים לעומס רגשי. מטרתי בעבודה זו הינה זיהוי הדרישות שאירועים בודדים במהלך עבודה של עובדים אשר מייצרים עומס רגשי. הנחת העבודה היא כי חשיפה של עובדים לאירועים אלה עשויה ללמד באופן אובייקטיבי על חוויית העומס הרגשי של אותם עובדים.

המחקר נעשה בהנחייתן של פרופ׳ ענת רפאלי ופרופ״ח גלית יום-טוב בפקולטה להנדסת תעשייה וניהול ע״ש וויליאם דוידסון.

אני מודה לטכניון, קרן גוטווירט, וקרן משפחת קראון על התמיכה הכספית הנדיבה בהשתלמותי.

עומס רגשי במערכות שירות: הגדרה ובחינה של עומס רגשי על ביצועי עובדים

חיבור על מחקר

לשם מילוי חלקי של הדרישות לקבלת תואר דוקטור לפילוסופיה

דניאל אלטמן

הוגש לסנט הטכניון - מכון טכנולוגי לישראל

ניסן, התשפ״א, מרץ 2021