



# Initial elevation bias in subjective reports

Patrick E. Shrout<sup>a,1</sup>, Gertraud Stadler<sup>b</sup>, Sean P. Lane<sup>c</sup>, M. Joy McClure<sup>d</sup>, Grace L. Jackson<sup>e</sup>, Frederick D. Clavé<sup>f</sup>, Masumi Iida<sup>g</sup>, Marci E. J. Gleason<sup>h</sup>, Joy H. Xu<sup>a</sup>, and Niall Bolger<sup>i</sup>

<sup>a</sup>Department of Psychology, New York University, New York, NY 10003; <sup>b</sup>Aberdeen Health Psychology Group, Department of Applied Health Sciences, University of Aberdeen, Aberdeen, AB25 2ZD, Scotland, United Kingdom; <sup>c</sup>Department of Psychological Sciences, Purdue University, West Lafayette, IN 47906; <sup>d</sup>Gordon F. Derner School of Psychology, Adelphi University, Garden City, NY 11530; <sup>e</sup>Department of Psychology, University of California, Los Angeles, CA 90095; <sup>f</sup>Department of Psychology, Iowa State University, Ames, IA 50011; <sup>g</sup>T. Denny Sanford School of Social and Family Dynamics, Arizona State University, Tempe, AZ 85287; <sup>h</sup>Department of Human Development and Family Science, The University of Texas at Austin, Austin, TX 78712; and <sup>i</sup>Department of Psychology, Columbia University, New York, NY 10027

Edited by Susan T. Fiske, Princeton University, Princeton, NJ, and approved November 20, 2017 (received for review July 10, 2017)

**People's reports of their thoughts, feelings, and behaviors are used in many fields of biomedical and social science. When these states have been studied over time, researchers have often observed an unpredicted and puzzling decrease with repeated assessment. When noted, this pattern has been called an "attenuation effect," suggesting that the effect is due to bias in later reports. However, the pattern could also be consistent with an initial elevation bias. We present systematic, experimental investigations of this effect in four field studies (study 1:  $n = 870$ ; study 2:  $n = 246$ ; study 3:  $n = 870$ ; study 4:  $n = 141$ ). Findings show clear support for an initial elevation bias rather than a later decline. This bias is larger for reports of internal states than for behaviors and for negative mental states and physical symptoms than for positive states. We encourage increased awareness and investigation of this initial elevation bias in all research using subjective reports.**

surveys | longitudinal design | response bias | epidemiology | subjective reports

People's reports of their thoughts, feelings, and behaviors are used in many fields of biomedical and social science as well as in clinical practice. Epidemiologists, for example, use symptom reports in studies of disease; sociologists and economists use social survey reports to study economic and social behavior; and neuroscientists use reports of emotional states to understand patterns of brain activation. Often participants are asked to provide reports at multiple points in time, and investigators use these to test predictions about causal processes and to describe developmental change.

However, at least in the case of negatively toned states such as mental and physical symptoms, concern has been expressed about the validity of repeated assessments. For over 50 y, researchers have noticed that reports of levels and severity of certain symptoms and maladaptive experience diminish over repeated assessments (1–4).

In a commentary on the National Institute of Mental Health Epidemiological Catchment Area study results, Robins noted that "... in second interviews respondents frequently fail to report [lifetime] symptoms that they reported in the first interview. The strange result is that the proportion of life-time cases at reinterview seems to shrink if the second interview rather than the first is used to calculate prevalences" (2). This pattern of decrease after initial report in longitudinal studies has been called the "attenuation effect" (5–7), suggesting that the pattern is due to a decline in later reports.

This label is premature in our view because previous evidence cannot distinguish whether the decline is due to upward bias in initial reports, to downward bias in subsequent reports, or to a combination of both. Upward bias would occur if, at the initial survey, respondents report higher levels or severity of the target state than they actually experience, and downward bias would occur if participants tend to report lower levels or severity than they actually experience. There is also the possibility that

the initial survey experience literally changed the course of the subjective process. Called "measurement reactivity" (8), the subsequent reports are not necessarily biased, but they would not reflect the longitudinal process in the part of the population that was not assessed. Distinguishing bias in the initial assessment or subsequent assessments from measurement reactivity is challenging. We attempt to do so in this article, but we first document the phenomenon that the level and severity of subjective reports often decline over repeated assessments. We began our work using a neutral label for the typical pattern: the "initial elevation or later decline" (IELD) effect.

We present the results of four field experiments. The first, second, and fourth experiments used intensive longitudinal data collection with daily diaries to examine moods, symptoms, and study habits before stressful examinations. The third experiment used a bimonthly panel survey design to examine these outcomes in the broader context of freshmen's adjustment to college. These target outcomes were selected to allow a contrast between more subjective reports (e.g., moods) and more objective behavior (e.g., study time). They also allow a contrast between negative (e.g., physical symptoms) and positive (e.g., vigor) states, because the previous literature has tended to report the IELD effect primarily on negative outcomes. Previous results are also possibly confounded by naturally occurring events or time-period variation. Each of our experiments removed this important confound by disentangling the timing of the report (e.g., relative to an examination) from the serial position of the report within a longitudinal series. Participants were randomly assigned to different starting times, allowing an experimental examination of the IELD effect.

## Significance

**People's reports of their own thoughts, feelings, and behaviors are essential assessment tools in biomedical and social science. When subjective states have been studied over time, however, researchers have often observed an unpredicted and puzzling decrease with repeated assessments. Our results across multiple outcomes in four field experiments suggest that this pattern is due to an initial elevation bias. This effect is larger for reports of internal states than for behaviors and for negative mental states and physical symptoms than for positive states. This initial elevation bias needs to be considered in all types of research using subjective reports.**

Author contributions: P.E.S. and N.B. designed research; P.E.S., G.S., S.P.L., M.J.M., G.L.J., F.D.C., M.I., M.E.J.G., J.H.X., and N.B. performed research; P.E.S., G.S., S.P.L., M.I., M.E.J.G., and J.H.X. analyzed data; and P.E.S., G.S., S.P.L., M.J.M., G.L.J., F.D.C., M.I., M.E.J.G., J.H.X., and N.B. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Published under the PNAS license.

<sup>1</sup>To whom correspondence should be addressed. Email: pat.shrout@nyu.edu.

Across all four experimental studies and across multiple outcomes, we find a robust IELD pattern. Overall, the findings converge to support an initial elevation bias rather than a later decline. The standardized effects range from small to medium, with larger and more reliable effects for reports on internal states than for behaviors and for reports on negative states and physical symptoms than for positive states. The effects were also larger and more reliable in the daily diary design than in the panel design and when we compared persons with other persons starting at different times rather than with themselves over time.

The third and fourth field experiments were also designed as preliminary tests of two potential mechanisms of the IELD effect. The first mechanism is a phenomenological process whereby the experience of assessment drives real changes in participants' internal states. This process could encompass initial elevation, where increased self-awareness exacerbates emotional intensity (7, 9), a later decline through a therapeutic process, whereby expressing distress causes actual declines in subsequent distress and other negative outcomes (10, 11), or both. This was examined in the third field experiment by contrasting the reports target persons made about themselves with reports roommates made about the targets. Evidence was inconsistent with either effect because the initial elevation and later decline pattern was present for both self- and roommate reports.

The second mechanism involves conversational norms, whereby participants interpret each repeated assessment as part of an ongoing conversation. This means that, irrespective of the exact questions asked of them, participants respond in ways that are normative in a conversational context. These norms include (i) providing only what participants perceive to be relevant information and (ii) in later assessments, providing only what they perceive to be new information (12, 13). The former could lead to an initial elevation bias (where participants provide exaggerated reports of their current states because they include information from earlier or future time periods), and the latter could lead to a subsequent decline (by participants ignoring their actual current states and only reporting what is new about those states).

In the fourth field experiment, participants were randomly assigned to experience either a single 2-wk diary study about stress approaching an examination (i.e., one conversation) or to experience two sequential but ostensibly unrelated 1-wk diary studies, the first about health and the second about the approaching examination (i.e., two conversations). Comparisons between the one-study and two-study participants did not support the conversational norms mechanism. We note that these are only preliminary tests of these mechanisms, and further research is necessary. Moreover, these are only two potential mechanisms; other mechanisms are discussed below.

That said, we can state that the IELD effect is robust, appearing for multiple outcomes across four field experiments. It is evident in both diary and panel designs, in the context of acute or chronic stress, and for reports about the self and about a familiar other. We judge that the IELD effect is due to initial elevation bias and that it can occur in all types of research that use subjective reports. This claim and the broad implications are discussed further below.

### Study 1: Demonstrating the IELD Effect in a Diary Study of Bar Examination Preparation

Study 1 was a survey of recent law school graduates and their intimate partners carried out between 2001–2003 during the 4-wk period before the graduates took a state bar admissions examination and the week following the examination. The primary goal of the survey was to study support processes (14), but the survey also had a design component to study the IELD effect. Couples were assigned to one of four conditions. The majority of couples ( $n = 393$ ) were in condition 1: They completed twice-daily reports (morning and evening, hereafter “AM” and “PM,

respectively) for 44 d, starting 35 d before the examination, including the two examination days, and continuing for 7 d afterward. The twice-daily design allowed the initial report to be studied in absolute terms (first AM report) and more relative terms (reports in first day of study). In conditions 2, 3, and 4, groups of couples were asked to make initial reports on day 22 (2 wk before the examination;  $n = 66$ ), on day 35 (1 d before the examination;  $n = 27$ ), or on day 44 (1 wk after the examination;  $n = 34$ ). We examine IELD patterns for reports of anxious and vigorous mood, physical symptoms, and time spent studying (examinees only), using both within-subjects comparisons of respondents' initial reports (day 1 of reporting) with matched subsequent reports (day 8 of reporting) and between-subjects comparisons of persons in condition 1 (the daily diary) with persons responding for the first time closer to the examination (conditions 2–4). Fig. 1 illustrates the design and results for PM anxious mood reports for the examinees.

The examination was on days 36 and 37. The survey form for all four conditions was the same; it obtained “right now” reports of moods and retrospective reports of physical symptoms and time spent studying during the prior 24 h. One can see from Fig. 1 (group 1) that PM anxiety generally increased from the first week of examination preparation to the days of the examination and that there was a remarkable recovery to low daily anxiety in the week following the examination. This pattern is completely consistent with theoretical and intuitive expectations of the effect of stressor temporal imminence (15) and supports a claim of construct validity for the daily measure. We argue that this pattern allows us to consider the group 1 diary reports on days 22, 35, and 44 as a standard that allows the bias of the initial reports on those days to be estimated. We also use this pattern to examine a more conservative estimate of bias for the first diary day of the series. Stressor imminence predicts that days closer to the examination will be more distressing than days more distal to the examination. Anxiety on day 8, which is the same day of the week as day 1, should therefore be higher on average than on day 1. Insofar as the opposite is observed, an IELD effect can be conservatively estimated.

The day 1 minus day 8 comparison for AM anxiety yielded a Cohen's  $d$  of 0.47 [ $t(285) = 8.28, P < 0.0001$ ]. (We report IELD effects as if they are initial elevations rather than later declines. This allows us to represent the expected effects as positive in sign rather than negative.) The comparison for PM anxiety yielded a  $d$  of 0.26 [ $t(284) = 4.50, P < 0.0001$ ]. Cohen's  $d$  (ref. 10, p. 20) estimates were obtained by calculating the mean within-subject change and dividing it by the pooled between-subject SD. According to Cohen, small, medium, and large effect sizes for  $d$  are 0.2, 0.5, and 0.8, respectively; the AM anxiety effect size would thus be classified as medium and the PM anxiety effect size as small.

Table 1 shows the results for the within-subject IELD effects for anxious mood as well as for AM and PM vigorous mood, PM physical symptoms, and PM reports of how many hours the examinee spent studying. Like anxious mood, there was a significant IELD effect for reports of physical symptoms [ $d = 0.39, t(288) = 5.28, P < 0.0001$ ] but not for vigorous mood in the AM or PM, or for study time.

We next compared the anxious mood reports of the diary group at days 22, 35, and 44 with the reports of the three panels that completed the same survey form for the first time on those days. The between-subjects comparisons at day 22 revealed significant IELD effects for the negative experiences of anxious moods and physical symptoms but not for vigorous mood or study time. The effect sizes of the significant IELD effects tended to be larger for the between-subjects tests than for within-subject tests. In Cohen's classification, the AM anxious mood effect [ $d = 0.82, t(332) = 5.46, P < 0.0001$ ] and the PM physical symptoms effect [ $d = 0.80, t(332) = 5.72, P < 0.0001$ ] were large,

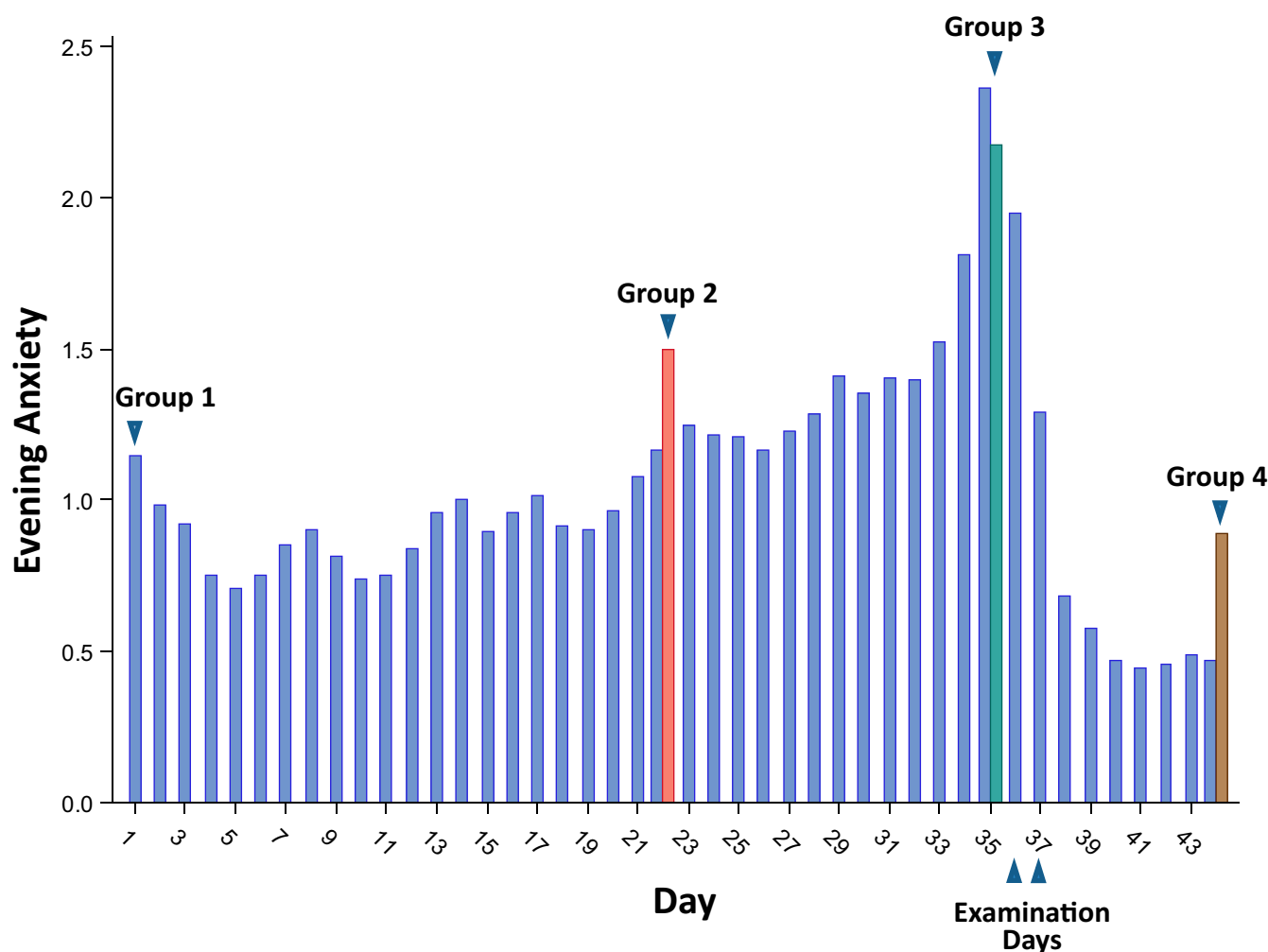


Fig. 1. PM anxiety in study 1 over 44 d around the 2-d state bar examination (on d 36 and 37) for examinees in four experimental conditions with different assessment start dates.

and the PM anxious mood effect was small to medium [ $d = 0.35$ ,  $t(331) = 2.29$ ,  $P < 0.03$ ].

On day 35, the eve of the examination, only one outcome showed an IELD effect: Those in the diary group reported fewer physical symptoms than those in day 35-only group. As is apparent in Fig. 1, there were no group differences in PM anxiety. In contrast, on day 44 there were large IELD effects for AM anxious mood [ $d = 0.84$ ,  $t(263) = 3.83$ ,  $P < 0.001$ ] and physical symptoms [ $d = 0.88$ ,  $t(266) = 4.32$ ,  $P < 0.001$ ] and a medium effect for PM anxious mood [ $d = 0.58$ ,  $t(263) = 2.62$ ,  $P < 0.001$ ]. Even though the participants in condition 4 who were responding to the survey were typically on vacation following the examination, they reported elevated current anxious mood and physical symptoms in their first experience with the study survey. This finding is perhaps the most compelling reason from study 1 to conclude that the IELD pattern is due to an initial elevation and not to a later decline.

To check if the IELD effects were specific to examinees, i.e., only the member of each couple who was directly exposed to the stressor, we looked at the analogous pattern in the partner. Partners were in the same condition (diary or one of the panels) as their examinee. Table 1 shows that, like the examinees, the partners tended to report more anxious mood and physical symptoms on their first day relative to day 8, but they also reported a slightly elevated level of vigorous mood in the evening

[ $d = 0.18$ ,  $t(268) = 2.56$ ,  $P < 0.001$ ] on day 1 versus day 8. The between-subjects contrasts at days 22, 35, and 44 for the partners produced consistent results: Comparing partners in panels to the relevant days for partners in the diary condition, panelist-partners reported higher levels of anxious moods in the AM and PM as well as physical symptoms. These effects were typically large in magnitude, and it is noteworthy that they were observed even on the day before the examination, a point in time where a comparison of examinee outcomes showed almost no differences. However, we found no IELD effects for vigorous mood in any of the between-subjects comparisons for partners.

In sum, study 1 showed that IELD effects for daily reports of mood, physical symptoms, and study time were as large as  $d = 0.88$  for examinees and  $d = 1.09$  for partners, with median effect sizes of 0.26 (examinees) and 0.24 (partners) across all variables and times. Effect sizes for subjective negative internal states and physical symptoms were larger (median effect of 0.54) than those for the more concrete behavior of study time (median effect 0.00) and for positive states (median effect 0.07). Further, these results support an initial elevation bias rather than a later decline: As the examination approaches, participants are typically more anxious, but for group 1, the within-person comparison shows a decrease in anxiety from day 1 to day 8. Groups 2 and 4 both show a distinct elevation on their starting day relative to group 1; this is particularly noteworthy for group 4, who on day

**Table 1. Estimates of IELD effects for examinees and partners based on comparison of diary days (within person) and of diary and panel groups (between persons) in study 1**

Reporter	Within person comparisons of days 1 and 8			Panel vs. diary conditions at day 22			Panel vs. diary conditions at day 35			Panel vs. diary conditions at day 44						
	Effect size	LB	UB	Effect size	LB	UB	Effect size	LB	UB	Effect size	LB	UB				
<b>Examinee</b>																
Anxious mood AM	0.47	***	0.36	0.58	0.82	***	0.52	1.12	0.31	-0.28	0.89	0.84	***	0.41	1.27	
Anxious mood PM	0.26	***	0.15	0.37	0.35	*	0.05	0.66	-0.20	-0.79	0.40	0.58	**	0.14	1.01	
Vigorous mood AM	0.09		-0.03	0.21	0.07		-0.20	0.34	0.44	+	-0.01	0.89	-0.17	-0.60	0.26	
Vigorous mood PM	0.10	+	-0.01	0.21	0.01		-0.25	0.27	0.06		-0.40	0.52	-0.20	-0.63	0.23	
Physical symptoms PM	0.39	***	0.25	0.54	0.80	***	0.53	1.08	0.79	***	0.21	1.37	0.88	***	0.48	1.27
Study time PM	-0.02		-0.13	0.09	0.01		-0.26	0.28	-0.47	+	-1.01	0.08	0.26	-0.24	0.76	
<b>Partner</b>																
Anxious mood AM	0.47	***	0.33	0.61	0.96	***	0.68	1.25	1.09	***	0.57	1.61	0.49	*	0.07	0.91
Anxious mood PM	0.14	*	0.00	0.29	0.64	***	0.34	0.94	0.64	*	0.05	1.22	0.55	**	0.14	0.97
Vigorous mood AM	0.10		-0.03	0.23	0.20		-0.09	0.49	0.24		-0.21	0.68	-0.03	-0.45	0.40	
Vigorous mood PM	0.18	*	0.04	0.31	0.11		-0.18	0.40	0.24		-0.21	0.69	-0.35	-0.77	0.08	
Physical symptom PM	0.50	***	0.33	0.66	0.92	***	0.60	1.24	0.60	**	0.15	1.05	0.64	***	0.25	1.03

Notes: Effect sizes are in Cohen's *d* metric. Significance levels: + $P < 0.10$ ; \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ . LB and UB are lower and upper bounds of 95% CIs, respectively. *Dfs* for one-sample *t* tests of within-person comparisons ranged from 271–288. *Dfs* for two-sample *t* tests of between-group comparisons ranged from 312–332 for the day 22 comparison, from 237–278 for the day 35 comparison, and from 258–263 for the day 33 condition.

44 are reporting for the first time during the recovery period after the examination, when anxiety should be low. Both examinees and partners who make their initial report on day 44 report levels of anxiety and physical symptoms that are remarkably elevated relative to what is expected.

### Study 2: Demonstrating the IELD Effect in a Diary Study of College Examination Preparation

Study 2 conceptually replicates study 1, focusing on a shorter period of time before a less critical examination and varying the start dates in the diary design. It did not include partners or any start dates after the examination. Participants were college students preparing for premedical science examinations between the fall semesters of 2010 and 2011. They were randomly assigned to start the diary survey on one of seven different days (ranging from 9 d to 2 d before the examination). Each participant, irrespective of when they started, was required to complete diaries for 14 consecutive days. They reported mood at waking, and at bedtime they reported mood, physical symptoms, and the amount of time spent studying. Fig. 2 shows how the average anxious mood (*Upper Left*) and physical symptom count (*Lower Left*) varied over days in each of the seven groups. Again, if there were no IELD effects anxiety, would be expected to increase until the examination and then drop afterward. Ignoring the initial responses, the expected pattern is observed. However, the initial responses reveal a pattern that is more consistent with an initial elevation bias rather than a later decline: One can see that the initial report tended to be higher than the adjacent reports across all seven groups, before coming back to what is presumably the phenomenologically real change (e.g., an increase in anxiety) as the examination approaches.

Table 2 shows the estimated IELD effects for the AM and PM anxious mood and vigorous mood and PM physical symptoms and study time. These estimates are based on comparisons of the first reports of persons in groups 2 through 7 with the reports on the same day of participants who had started the diary process earlier. There was a medium effect for AM anxiety [ $d = 0.56$ ,  $t(225) = 5.98$ ,  $P < 0.0001$ ] but no evidence of an IELD effect for PM anxiety on the same day. In contrast, there was no evidence of an IELD effect for AM vigor, but there was a small-to-medium effect for vigor reported in the evening [ $d = 0.38$ ,  $t(225) = 3.27$ ,

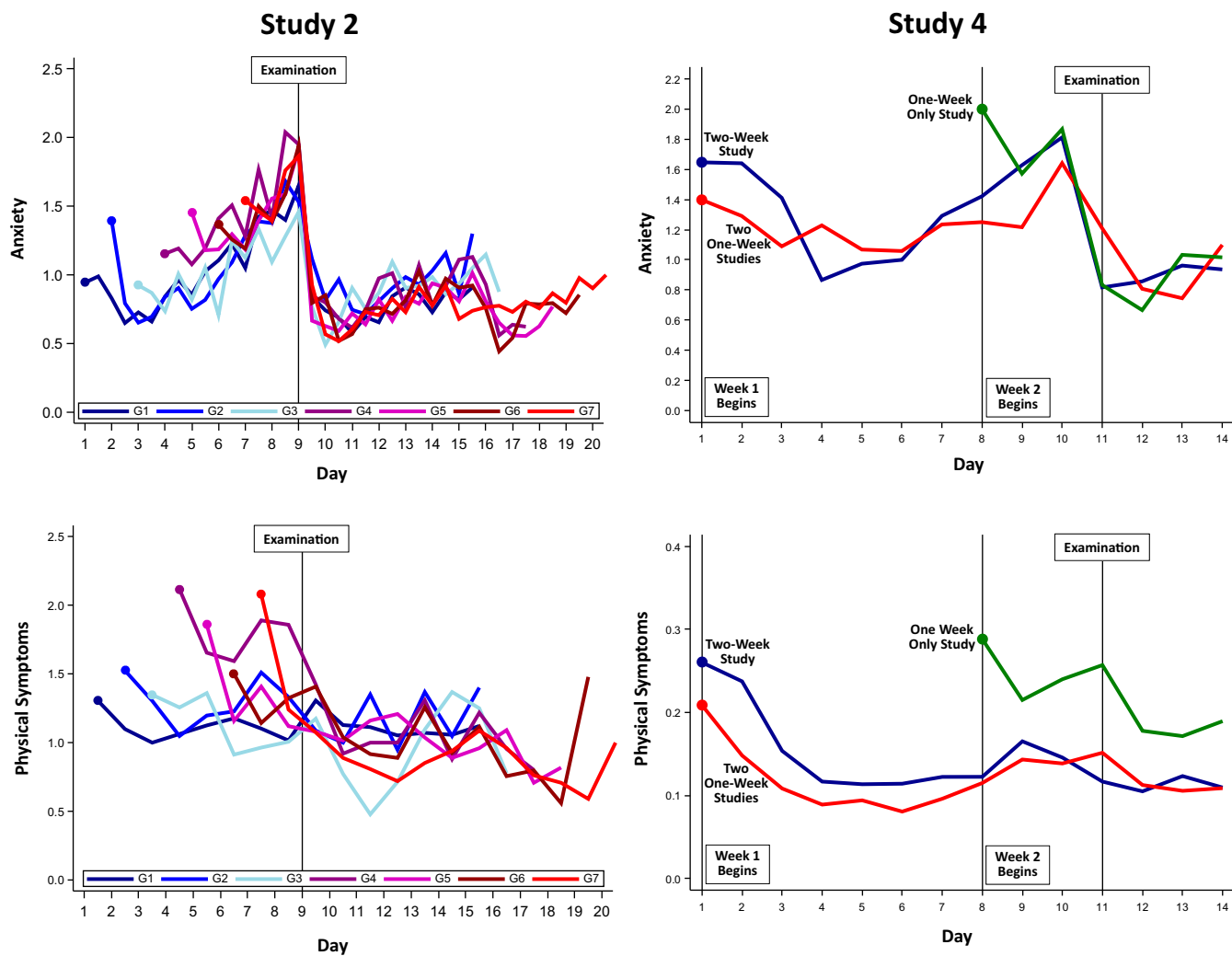
$P < 0.002$ ]. There was also evidence of a small-to-medium effect for physical symptoms, which were reported only in the evening [ $d = 0.34$ ,  $t(225) = 4.15$ ,  $P < 0.0001$ ], and also for PM reports of study time [ $d = 0.43$ ,  $t(225) = 4.36$ ,  $P < 0.0001$ ].

Study 2 results can be compared with the between-person examinee results from study 1. For AM anxious mood, the IELD effect was replicated, although it was smaller than the average effect reported in study 1; for the PM anxious mood effect there was no IELD replication. In addition, study 2 found IELD effects that were not apparent in study 1: positive mood (vigor) and the more objective report of study time.

To what extent does the IELD effect confound inferences about the size of temporal effects? The data from study 2 provide a clue. Fig. 2, *Upper Left* shows a sharp increase in anxiety as the examination draws near that drops dramatically after the examination. From the day before to the day after the examination the average drop in AM anxiety is 1.00 (SE = 0.12) in standardized effect units. Had the day before been the initial survey, we would expect from Table 2 that the change would be 0.56 larger: 1.56 rather than 1.00 (i.e., 1.56 times too large). In contrast, the change in reported PM physical symptoms from the day before to the day after the examination (Fig. 2, *Lower Left*) is 0.18 (SE = 0.08). The IELD effect from Table 2 is 0.34. Had the first measure been biased by the IELD effect, the estimated effect of the examination would have been nearly three times (2.89) too large. We conclude from this exercise that the impact of the IELD effect on substantive findings is not constant and needs to be considered in each context.

These first two studies were both limited to the context of upcoming stressful events and to the use of diary designs, with reports made on sequential days about recent or current events and feelings. Is the IELD effect limited to this context and design? In study 3 we examined whether IELD effects are found in a milder chronic stress context (everyday college life) rather than in an acute stressor context (examination preparation). This study uses bimonthly measurement over 8 mo with a design that reduces the confounding of possible IELD effects with period effects (e.g., the beginning of fall semester, Thanksgiving) from first-interview effects.

Further, study 3 sought to specifically examine a potential phenomenological mechanism for the IELD effect. That is, the IELD may not result from a bias in reporting but from a phenomenologically real, albeit measurement-driven, change in state. Previous



**Fig. 2.** Anxiety (*Upper*) and symptoms (*Lower*) in studies 2 (*Left*) and 4 (*Right*) over the days of the study in the experimental conditions with different assessment start dates.

researchers (8) have proposed that the IELD may be observed for negative states, not because of a bias in reporting but because of genuine ameliorative effects of self-disclosure and study participation leading to later decline. Alternatively, participants beginning the study may experience increased self-awareness, which could increase emotional intensity (9), contributing to initial elevation. To examine this possibility, study 3 includes surveys of a college roommate's perception of the other roommate's level of distress. For roommates providing repeated reports about another person instead of the self, a phenomenological effect would not be plausible.

### Study 3: Demonstrating the IELD Effect in a Panel Study of College Roommates and Testing a Therapeutic Mechanism

Study 3 participants were undergraduate students ( $n = 870$ ) who were recruited in the 2009–2010 academic year for a study of their experiences of “college life”; the timing and content of the surveys were unconnected with specific stressors. Eighty-five percent of the participants were recruited as roommate pairs; half of these pairs were randomly chosen to report on their own college experience and the others to report on their roommate's experience. Repeated online surveys were scheduled for October, December, February, and April, with participants randomly assigned to groups that began in October ( $n = 171$  self-report;  $n = 124$  roommate report), in

December ( $n = 158$  self-report;  $n = 127$  roommate report), or in February ( $n = 162$  self-report;  $n = 128$  roommate report). Self-report participants revealed their current anxious and vigorous mood, 6-wk recall of physical symptoms, and 6-wk recall of mental distress symptoms (K10 scale) (16). Roommate-reporting participants rated mental distress symptoms for their college roommates.

**Table 2.** Study 2 IELD effect estimates for diary college student's self-reports before science examination

Outcome	Effect	LB	UB
Anxiety AM	0.56***	0.38	0.75
Anxiety PM	0.10	−0.10	0.30
Vigor AM	0.11	−0.10	0.33
Vigor PM	0.38**	0.15	0.60
Physical symptoms PM	0.34***	0.18	0.50
Study time PM	0.43***	0.24	0.63

Effect sizes are in Cohen's  $d$  metric and were estimated in a generalized linear model that compared first reports with the average of reports of participants whose report was not their first report. The model adjusted for position of day relative to examination and for correlated residuals. Significance levels: \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ . LB and UB are lower and upper bounds of 95% CIs, respectively.  $df$  for tests, 223.

**Table 3. Study 3 IELD effect estimates for college students' self-reports and for reports on roommates**

Outcome	Effect size	LB	UB
Participant self-report			
Anxious mood (current)	0.16***	0.05	0.27
Vigorous mood (current)	0.15**	0.05	0.26
Physical symptoms (6 wk)	0.18***	0.11	0.26
K10 mental distress (6 wk)	0.29***	0.21	0.37
Participant report on roommate			
K10 mental distress (6 wk)	0.13***	0.03	0.23

Effect sizes are in Cohen's  $d$  metric and were estimated in a linear mixed model that compared first reports with the average of reports of participants whose report was not their first report. The model adjusted for early/late time-points in the semester and random intercepts. Significance levels: \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ . LB and UB are lower and upper bounds of 95% CIs, respectively. The participant results were based on  $n = 455$ , and the roommate results were based on  $n = 345$ .

Table 3 shows estimates of the IELD effects for the self and roommate reports. The effect estimates, which are based on repeated-between-person comparisons, were obtained using a mixed-effects model that included an indicator of the initial assessment, an adjustment for periods close to scheduled examinations (December and April), and a person random intercept. All the IELD estimates for self-reports were significantly different from zero, but the magnitude of the effects was small (ranging from  $d = 0.15$  for current vigor to  $d = 0.29$  for own mental distress). Taken together, studies 1 through 3 showed that IELD effects were omnipresent for self-reports of negative states and physical symptoms, with smaller, less robust effects for positive states. The effect was smaller in this study than in the previous two; this could be attributable to the change in design from diary to panel, the change in context from acute to chronic stress, or a combination of both.

We were especially interested in whether the IELD would be present for participants' reports of their roommates' mental distress. If the IELD is driven by a phenomenological effect, wherein one's experience of measurement drives an actual change in one's internal state, then the IELD would not be present when reporting about another person. Contrary to this hypothesis, there was a significant IELD effect for mental distress reports on roommates [ $d = 0.13$ ,  $t(341) = 2.45$ ,  $P < 0.015$ ].

Another explanation of the IELD effect that remains plausible is the conversational norm mechanism, whereby participants interpret the repeated measurements as part of an ongoing conversation and, in an effort to follow norms, focus on providing relevant information and on updating previous information (12). For example, in study 1 participants were recruited from law schools to report on their bar examination preparation. One group was randomly selected to give their first report after the examination had taken place. Although they were likely to have recovered from the stress of the examination held 7 d before, they might have assumed that the investigators were interested in how stressful the examination had been, even though the questions were phrased for current symptoms and problems. Study 2 was also framed around stressful experiences; participants' initial reports of negative states therefore may have been elevated by mixing their current state with their global sense of anxiety about the stressor. Although study 3 was not framed about a specific stressful examination, it was announced to be about "college life," which many students at competitive private universities might associate generally with stress.

The design of the first three studies confounded experience with the measures with the framing of what the researcher might want to know. Moreover, participants' first reports were also their first experience completing the measures. Study 4 was designed to test the conversational norms mechanism by sepa-

rating experience with the measures from the initiation of the conversation with the researcher.

#### Study 4: Testing the Conversational Norms Mechanism for IELD in a Single Study Versus Sequential Studies Design

The final study, like study 2, again focused on undergraduates preparing for a difficult examination. All participants were initially recruited (in 2011–2012) to participate in a study that was called the "Examination Preparation Study." They were randomly assigned to two groups: The participants in group 1 ( $n = 53$ ) were simply asked to complete 14 d of diaries for the Examination Preparation Study (identical to group 1 in study 2), whereas participants in group 2 ( $n = 66$ ) were given a more complicated story. On the night before the study was to begin, they were told that the Examination Preparation Study had reached its quota of subjects and that they would not be needed until a week later. They were then told that, if they were interested, they could participate in the interim in a 1-wk diary study that paid the same amount but that was on health run by a different faculty investigator. They were told that the College Health Study focused on everyday health behaviors in college students and was not at all concerned with the upcoming examination. Two thirds of those in group 2, a subgroup we call "group 2a" ( $n = 44$ ), agreed to be in the health study. They gave new informed consent for an ostensibly new faculty investigator and completed a weeklong diary study. Participants in the remaining subgroup, which we call "group 2b" ( $n = 22$ ), chose not to participate in the health study, but all of them chose to begin the Examination Preparation Study a week later. Thus, we had three groups: Group 1 completed diaries on 14 consecutive days, from 11 d before to 2 d after their examination; group 2a completed diaries for the unrelated health study for 7 d, switched to completing diaries for the Examination Preparation Study on day 8, and continued until day 14; group 2b enrolled in the Examination Preparation Study but did not begin completing diaries until day 8 and then continued to day 14 (identical to group 1 in study 2). These three groups are shown in Fig. 2, *Right*.

If the conversational norm mechanism were operating, on day 8, participants in both group 2a and group 2b would show an elevation compared with group 1: For group 2a and 2b, this day marked the start of a new conversation, whereas for group 1 this day merely continued the conversation they had already been having. Table 4 and Fig. 2, *Right* show that there were no between-persons IELD effects on day 8 for group 2a compared with group 1. There were, however, IELD effects for physical symptoms in within-subjects comparisons of day 1 with day 8 for group 1 and for both anxious mood and physical symptoms when group 2b was compared with group 1. Again, consistent with studies 1 and 2, the effects appear to be initial elevations rather than later declines: The approaching examination would not be consistent with a reduction in anxiety and physical symptoms. The first report for group 2b is elevated compared with groups 1 and 2a in a deviation from the presumably phenomenologically real increase associated with the examination.

#### Discussion

Previous researchers who noticed the unexpected and puzzling pattern of decline after initial assessments have labeled the pattern an "attenuation effect," implying an artifactual change (i.e., later decline) in later reports (5, 6). However, little systematic research about the size, scope, or basis of the effect has been done. We addressed this gap in four field experiments.

Overall, we find "attenuation" to be a misnomer: Our findings are generally consistent with an initial elevation bias, not with a later decline. This is evident within persons, with reports of anxiety initially decreasing despite an approaching examination in studies 1, 2, and 4. It is also strikingly evident between persons: Those participants who gave their first reports of anxiety 10 d after the

**Table 4. Results from study 4: One- vs. two-study design**

Outcome	Between group 2a vs. group 1			Within group 1: Day 8 vs. day 1			Between group 2b vs. group 1		
	Effect	LB	UB	Effect	LB	UB	Effect	LB	UB
Anxiety PM	−0.17	−0.60	0.27	0.17	−0.16	0.50	0.56*	0.00	1.12
Vigor PM	−0.11	−0.53	0.32	−0.09	−0.35	0.16	−0.02	−0.55	0.51
Physical symptoms PM	−0.04	−0.41	0.33	0.77***	0.45	1.09	0.92***	0.45	1.39
Study time PM	0.00	−0.45	0.45				−0.21	−0.77	0.34

Effect sizes are in Cohen's *d* metric. Significance levels: \* $P < 0.05$ ; \*\*\* $P < 0.001$ . LB and UB are lower and upper bounds of 95% CIs, respectively. Group 1 completed a single examination preparation study for 14 d. Group 2a completed 1 wk of study and 1 wk of examination preparation study. Group 2b completed only 7 d of examination preparation study. Comparisons of group 2a and 1 were on 87 *df* for symptoms and 83 for study time. Within-group *t* tests were on 46 *df*. Comparisons of group 2b and 1 were on 68 *df* for symptoms and 66 for study time.

bar examination showed marked elevation compared with those who had already provided reports. Also, as illustrated in Fig. 2 for both studies 2 and 4, initial reports for each group are elevated before subsequently converging into a coherent longitudinal pattern. Furthermore, the hypothesis that respondents were essentially receiving a treatment that ameliorated distress, leading to a later decline, was inconsistent with the results of study 3. In that study roommate's reports on each other showed the same IELD pattern, contrary to the therapeutic mechanism proposed for genuine later decline.

The overall IELD effect is robust, appearing across multiple outcomes in four studies. It appears across different study designs (intensive longitudinal vs. panel, reporting on the self vs. reporting on one's roommate, within-person and between-persons) and context (acute vs. chronic stress, for stressed persons vs. their romantic partners). The size of the effect also varied, with larger effects for internal states than for behaviors and for negative states and physical symptoms than for positive states. Nearly all the effect sizes we observed (median Cohen's *d* in studies 1, 2, and 3: 0.26, 0.34, and 0.16, respectively) are likely to be of practical significance, whether in establishing clinical cutoffs for depression or physical discomfort or simply in establishing benchmarks for responses to questions of emotions, physical symptoms, and a host of other subjective reports. We argue that the robust IELD findings are likely to be due to initial bias and therefore that initial bias effects need to be considered when interpreting survey data on subjective reports.

The present work has several important strengths: It presents experimental investigations of the pattern of decrease after initial reports in longitudinal research and systematically considers whether the pattern is a bias in initial elevation or later decline. We present a large body of evidence collected in four field experiments. The experimental design allows us to demonstrate the effect between-persons as well as within-persons, and, further, the between-person comparisons support our proposal that the process of starting the study causes the initial elevation bias. Finally, the replication of the initial elevation bias across each of these four field experiments eliminates potential artifactual explanations for the effect, such as an unanticipated secular event that affects reporting.

There are also limitations in the present work. First, our four experiments were carried out with students or law school graduates who were either facing scheduled examinations or engaged in an academic semester. We studied these participants because they were easily recruited and were willing to provide the intensive longitudinal data we required. Although they come from specialized populations, the IELD effects they displayed are consistent with similar effects that have been reported in epidemiological surveys (2, 17) of general populations. Similar drops in self-reported symptoms from the first to second wave of data continue to be reported in studies of psychopathology (18, 19) and health behavior (20). We hope that the magnitude of the

IELD effects will be studied as investigators become aware of the phenomenon.

A second limitation is the range of possible mechanisms considered. We proposed and tested two mechanisms, but further work is needed. Although we present paradigms for testing the phenomenological mechanism and the conversational norms mechanism, each of these merits a program of research beyond the scope of this paper. Our initial findings are inconsistent with these two mechanisms, but they should not yet be ruled out. Further, these are but two of a variety of potential mechanisms. One in particular that we believe to be promising is a learning mechanism, whereby increasing familiarity with a measure leads to less extreme reporting (6).

A similar mechanism—meaning-making—has been studied by Knowles and colleagues to explain an IELD effect at the micro-level, i.e., shifts in a single measurement session, within a single instrument, where participants' responses to items presented early are more extreme than items presented later (1). Although this pattern is consistent with IELD, it operates on a different scale from the one we have examined in the present research and so may operate by different psychological principles.

Other examples of IELD effects that differ from the present work would include those on longer timescales or with less subjectivity in measurement. We previously cited Robins (2) who noted that incidence rates of mental disorders cannot be accurately estimated from retrospective lifetime prevalence surveys that show the IELD pattern. The IELD effects of complex retrospective health surveys might be due to processes such as learning that endorsing screening questions leads to additional questions about the scope and severity of symptoms (6, 21). We sought to eliminate such additional mechanisms in the work reported here. Our work speaks primarily to repeated measurement of subjective reports on current internal states and recent behavior.

The implications of the initial elevation bias as demonstrated in our present work are nonetheless wide-ranging. A bias due to later decline would have been problematic for research using repeated measurements. However, an initial elevation bias potentially extends to all research using subjective reports, including cross-sectional designs. In such designs, the potentially biased first report is the only report. When social scientists study reactions to national or world events, or when medical researchers screen for disorders, elevation biases lead to false conclusions and to false-positive identifications of screened participants who in fact do not have the disorder. Randomized studies will be resilient to the bias, but when baseline measures are taken the IELD effect can contribute to placebo effects in the control condition. Although the IELD effect would not affect correlational studies if it were constant across individuals and variables, it could confound correlations if the magnitude varied with some characteristics of respondents. At the request of a reviewer, we checked to see if the IELD effect interacted with gender or majority/minority status in studies 1–3, but we found

no consistent effects for these variables. This does not mean that the IELD effect does not vary with other personality or communication characteristics. We urge survey researchers to add this effect to the list of other processes than can bias results, such as fatigue or acquiescence bias, and to look for possible bias in the survey results.

Recognizing an elevation bias leads to arguments in favor of designs that use repeated measurements, as they allow for the observation of and potential to adjust for the initial elevation bias, for example by dropping initial observations (e.g., ref. 22). Alternatively, researchers interested in subjective reports could consider providing prior experience with the given instrument before taking measurements of focal interest. This could be seen as analogous to common procedures with other types of measurement that require practice or establishing a baseline (e.g., computer tasks using reaction times, physiological recordings). We hope that the present paper will increase awareness of IELD effects generally and the possibility of initial elevation bias in particular.

## Materials and Methods

### Study 1: Establishing IELD in a Diary Study of the Bar Examination.

**Participants and design.** Participants were recent law school graduates who prepared for the state bar examination ( $n = 436$ , 55.5% female; mean age: 29.7 y) and their romantic partners ( $n = 434$ , 46.1% female; mean age: 30.2 y). Participants were recruited over 3 y (2001–2003) for a study of stress and support (14). They started 5 wk before the examination (preparation: days 1–35, examination: days 36 and 37) and continued for 1 wk afterward (recovery: days 38–44). In 2001 and 2002, couples were randomly assigned to one of four groups: daily diary from days 1–44, panels on days 22/35/44, or single assessments on day 35 only or on day 44 only. In 2003, all participants were assigned to the daily diary condition. For the current analyses we analyzed only day 22 in the second group. After eliminating eight persons with incomplete data, the sample sizes were 326 (daily diary), 60 (assessment on day 22), 19 (assessment on day 35), and 23 (assessment on day 44). Couples were paid \$150 in the daily diary condition and \$50 in the other three conditions; all participants were entered into a lottery to win \$1,000. This study and the following studies were approved by one or both of the New York University or Columbia University Institutional Review Boards, and informed consent was obtained from all participants in all studies.

**Measures.** For all studies in this article, we focused on one negative mood (anxiety), one positive mood (vigor), and physical symptoms. In studies 1, 3, and 4, examinees also reported their study time. In study 1, anxiety and vigor were measured using three items each from the Profile of Mood States (23). These have been shown to be reliable measures of between-person and within-person reports of mood (24). Participants reported their current mood twice, once upon awakening and once before going to bed. Response categories ranged from 1 (not at all) to 5 (extremely) (rescaled to a 0–4 range in the analysis). To assess physical symptoms, participants reported if any of four symptoms occurred in the previous 24 h: back or muscle ache, headache, upset stomach, and insomnia. These binary reports were averaged into a physical symptom index that ranged from 0 to 1. The index represents the overall burden of physical symptoms on a given day. To assess hours spent studying, the examinees were asked to report how many hours they spent studying for the examination. Reports ranged from 0 to 15, with an average of 8 on most days. The measures in study 1 were given as paper-and-pencil forms containing a total of 122 questions about mood, social support, stressors, and coping.

**Data analysis.** All five outcomes were divided by the average between-person SD so that differences can be interpreted as Cohen's  $d$  effect sizes (25). We estimated the within-person IELD effect by subtracting day 8 scores from day 1 scores and computing one-sample  $t$  tests on the differences as well as computing 95% confidence bounds. We used day 8 as the comparison score to adjust for day of the week effects. We estimated and tested between-persons IELD effects by comparing the responses of the participants who gave their first response on day 22, 35, or 44 with the responses of participants in the daily diary condition on the corresponding day. Independent sample  $t$  tests were used to test the statistical significance of the difference and to compute 95% confidence bounds. In all analyses we assume two-tailed tests with a type 1 error rate of 0.05. This initial study was part of a larger research program to understand stress, effect, and social processes in couples (14) and was conducted without a priori power analyses for the purpose of the IELD effect. The syntax and all the data used for study 1 analyses are available at <https://osf.io/8w2du/>.

### Study 2: IELD and Timing in Preparing for Premedical Examinations.

**Participants and design.** Participants were premed students ( $n = 246$ , median age = 20.03 y,  $SD = 2.08$  y, 70.5% female) recruited for a 2-wk diary study who received compensation for their time (up to \$50 and several chances to win \$250 lotteries). Participants who filled out at least one diary ( $n = 228$ ) were included in the analyses. Before the diary portion of the study all participants completed a background questionnaire. They were then randomly assigned to one of seven groups, each of which had a different start date ranging from 8 to 2 d before the examination. Participants were sent links to the online survey the day before they were to begin filling out the survey. Group 1 began the diary on day 8 before the examination and served as a comparison for the other six conditions. For this reason we allocated more participants to this group (using an 8:3 ratio) than to the other six groups to increase power for these contrasts. Participants in group 1 were also asked to complete 15 diary entries instead of 14. The remaining six groups filled out diaries over 14 d and were randomized to different starting days, ranging from 7 to 2 d before the examination. Each day participants completed two diaries, one within an hour of waking and one within an hour of going to bed in the evening. On average participants in the sample completed 13.61 AM diaries ( $SD = 1.02$ ) and 13.25 PM diaries ( $SD = 1.23$ ). Participants did not differ in terms of their demographic characteristics across groups.

**Measures.** Study 2 included measures of mood as assessed in study 1, an extended list of physical symptoms, and study time. The mood scales showed satisfactory reliability, with an average Cronbach's alpha of 0.80 for anxious mood and 0.71 for vigor, and average  $R_{change}$  of 0.82 and 0.69 for anxiety and vigor, respectively. To assess physical symptoms, each evening participants indicated whether they experienced any of eight different physical symptoms (nausea/upset stomach, sore throat, insomnia, constipation/diarrhea, headache, back/muscle ache, rash/irritation, runny nose/congestion), coded as "1" if present and "0" if not. Responses were summed to create an index of the total number of physical symptoms experienced on a given day. To assess study time, participants reported the number of hours they spent studying in the past 24 h in the PM diary.

**Data analysis.** We estimated the IELD effect by comparing the reports on the first day of reporting by groups 2 through 7 with the average of the reports of groups that had started the diary earlier. For example, the first report of group 2 was compared with the reports on the same day in group 1, and the first report of group 3 was compared with the average of the reports on that day from groups 1 and 2. We used the MIXED procedure of SAS to fit a general linear model that adjusted for average level of each day before the examination and adjusted for repeated measures by specifying a Toeplitz structure with six bands on the residual correlation matrix. A priori power analysis for this diary study was based on study 1 and a smaller study of repeatedly assessed depression by Sharpe and Gilbert (4) with an average effect size of  $d = 0.23$ . Using SAS PROC MIXED simulations to detect first day effects vs.  $d$  distant from the examination, we calculated an intended sample size of 260 participants with an estimated SE of 0.08 and 82% power to detect an effect of at least  $d = 0.23$ . The syntax and all the data used for study 2 analyses are available at <https://osf.io/jtdb8/>.

### Study 3: IELD in a Panel Study of College Experiences.

**Participants and design.** Undergraduates and their roommates were recruited from two urban private universities to participate in a longitudinal survey of college life over 8 mo [ $n = 870$ ; 742 (85.3%) were recruited as roommate pairs, mean age, 18.9 y ( $SD = 1.5$  y), 77% female]. Participants were randomly assigned to start the survey in October, December, or February. Roommate pairs were stratified in the random assignment so that the starting months were uncorrelated and balanced. All participants were asked to complete February and April assessments. The order of the self vs. roommate interview sections was randomized.

Of 870 persons recruited, 800 participants (self-report  $n = 455$ ; roommate report  $n = 345$ ) filled out at least one follow-up survey and were included in the analyses. Participants were asked to complete up to four bimonthly assessments describing their own or their roommate's (if enrolled in the study) psychological and physical health. For each pair of roommates, one target individual described herself/himself, and the roommate described the target individual. Participants received \$10 per survey completed, up to \$50 (background questionnaire and four bimonthly surveys), and had a chance to win one of five \$250 lotteries across the course of the study.

**Measures.** As in study 1, we investigated IELD in mood and physical symptoms and included overall distress. Mood was assessed as described in study 1. The average scores for anxiety and vigor were relatively low (anxiety mean: 0.78; vigor mean: 1.24). The reliability of anxiety and vigor was adequate,  $R = 0.77$  and 0.77, respectively. To assess physical symptoms, participants were asked to indicate the frequency of 14 physical symptoms over the past 6 wk on a scale ranging from 0 (never) to 3 (nearly every day). The physical symptoms



included headaches, asthma, cold/flu, nausea, and insomnia. The responses to the 14 items were averaged, with higher scores indicating more health symptoms. The average scores for physical symptoms was relatively low (mean = 0.43). The reliability of physical symptoms was also adequate,  $R = 0.85$ . We assessed distress with 10 items from a short distress scale (16) (an example of an item for reporting about self: "How often did you feel so nervous that nothing could calm you down?"; an example of an item for reporting about a roommate: "How often did your roommate feel so nervous that nothing could calm him/her down?"). Participants rated how they or the roommate had been feeling over the past 6 wk on a scale ranging from 0 (none of the time) to 4 (all of the time). The responses to the 10 items were averaged. The average scores for distress for participants reporting about themselves and those reporting about the roommate across experimental groups were relatively low (self: mean = 0.92; roommate: mean = 0.72). The reliability for self and roommate reports was  $R = 0.90$  in both groups.

**Data analysis.** We estimated the IELD effects by comparing the initial reports of participants who started the panel survey in December with those starting in October and comparing those starting in February with those who started in October or December. This was done with a multilevel model that included all available data; participants starting in October, December, and February could contribute respectively four, three, or two waves of data. The model adjusted for months near final examinations (December and April) and a random effect for the participant's average level of the outcome. The IELD effect itself was associated with a dummy variable that marked which survey point was first. We assumed smaller effect sizes in a panel study of the usual college experience with longer intervals than for a diary study of acute stress. We planned to recruit 810 participants, 270 per cell in a three-group (start dates October, December, February) experimental design to detect at least a small effect ( $d = 0.1$ ) for the contrast between first and second interviews with 81% power. The syntax and all of the data used for study 3 analyses are available at <https://osf.io/nr4yd/>.

#### Study 4: IELD and Conversational Norms.

**Participants and design.** Participants ( $n = 141$ , mean age: 18.8 y,  $SD = 1.4$  y, 73.9% female) were recruited for a 2-wk diary study before a major examination with the same methods as in study 2. Participants received compensation up to \$30 for their time. All participants were enrolled to complete 2 wk of a daily diary starting 10 d before the examination (days 1–10). After consent, they were randomly assigned to one of two conditions. In condition 1, participants completed the 2-wk diary as the Examination Preparation Study. Condition 2 was designed to induce students to participate in two seemingly separate, consecutive 1-wk diary studies run by two different research groups, in week 1 as the Daily Health Study and then in week 2 as the Examination Preparation Study, each with a different study purpose, research group, online survey platform, and survey layout. The 141 students who agreed to participate were randomly assigned to condition 1 or 2 (60 and 81 participants, respectively) at a ratio of 3:4 to compensate for expected drop-out in the two-study condition. Of the 60 assigned

to condition 1, 53 followed the link and provided usable data over the 2 wk. Of the 81 persons assigned to condition 2, 43 agreed to enroll in the Daily Health Study. One week later, these participants were invited to start the Examination Preparation Study. All those who enrolled in the Daily Health Study agreed to start the Examination Preparation Study in week 2. We call these 43 persons "condition 2a." Of the 38 participants who declined to participate in the Daily Health Study in week 1, 58% ( $n = 22$ ) enrolled in the 1-wk version of the Examination Preparation Study. We call these "condition 2b." Because self-selection undermined the initial randomization, we checked whether the three groups (conditions 1, 2a, and 2b) differed in age, gender, grade point average, or major. None of these comparisons revealed statistically significant differences.

Regardless of condition, participants completed identical items for the 2 wk (with additional health items in the Daily Health Study). After the 2-wk diary period, all participants completed a background questionnaire that, in addition to demographic information, assessed the extent to which participants in condition 2 were aware that the Daily Health Study and the Examination Preparation Study were actually identical diary surveys being run by one research team. Nine of the 43 persons (21%) in condition 2a suspected that the investigators of the Examination Preparation Study were the same as in the Daily Health Study. Eliminating these nine persons did not change the results. **Measures.** Measures in study 4 included mood, physical symptoms, and study time, as in study 2. To make the Daily Health Study in condition 2 more credible, we included more questions about health in study 4 than in studies 1–3. Participants answered a series of items that assessed the presence of a number of health issues over the course of the previous day, asking "How often during the past 24 h have you experienced the following health problems?" During week 1, participants in each condition responded to an extended checklist that included additional items assessing more differentiated headache symptoms (e.g., tension headache, migraine, sinus headache). However, the physical symptoms variable used in this analysis is based on the same eight items used in study 2.

**Data analysis.** To test the hypothesis that starting a new study would induce an IELD effect, we compared the level of responses on day 8 of group 1 (the 2-wk Examination Preparation study group) with the day 8 responses of group 2a (the group that completed 7 d of the health study and started a 1-wk Examination Preparation Study on day 8). The mean difference was tested with an independent sample  $t$  test. To estimate the within group presence of the IELD effect in group 1, we compared day 8 with day 1 and tested the difference with a paired  $t$  test. To test the IELD effect of group 2b, which started a 1-wk-long Examination Preparation Study without having completed the 1-wk health study, we compared their first day of reports with the day 8 reports of group 1 using an independent sample  $t$  test. Following guidance by Cohen (25), a sample size of 140 participants was calculated to detect a medium effect size of  $d = 0.5$  comparing the two experimental groups (the one-study condition vs. the two-study condition). The syntax and all the data used for study 4 analyses are available at <https://osf.io/y2r7s/>.

- Knowles ES, Coker MC, Scott RA, Cook DA, Neville JW (1996) Measurement-induced improvement in anxiety: Mean shifts with repeated assessment. *J Pers Soc Psychol* 71: 352–363.
- Robins LN (1985) Epidemiology: Reflections on testing the validity of psychiatric interviews. *Arch Gen Psychiatry* 42:918–924.
- Windle C (1954) Test-retest effect on personality questionnaires. *Educ Psychol Meas* 14:617–633.
- Sharpe JP, Gilbert DG (1998) Effects of repeated administration of the beck depression inventory and other measures of negative mood states. *Pers Individ Differ* 24:457–463.
- Jensen PS, Watanabe HK, Richters JE (1999) Who's up first? Testing for order effects in structured interviews using a counterbalanced experimental design. *J Abnorm Child Psychol* 27:439–445.
- Lucas CP, et al. (1999) Features of interviews questions associated with attenuation of symptom reports. *J Abnorm Child Psychol* 27:429–437.
- Piacentini J, et al. (1999) Informant-based determinants of symptom attenuation in structured child psychiatric interviews. *J Abnorm Child Psychol* 27:417–428.
- French DP, Sutton S (2010) Reactivity of measurement in health psychology: How much of a problem is it? What can be done about it? *Br J Health Psychol* 15:453–468.
- Silvia PJ (2002) Self-awareness and emotional intensity. *Cogn Emotion* 16:195–216.
- Neugebauer R, et al. (1992) Depressive symptoms in women in the six months after miscarriage. *Am J Obstet Gynecol* 166:104–109.
- Pennebaker JW (1997) Writing about emotional experiences as a therapeutic process. *Psychol Sci* 8:162–166.
- Schwarz N (1995) What respondents learn from questionnaires: The survey interview and the logic of conversation. *Int Stat Rev* 63:153–168.
- Edelbrock C, Crnic K, Bohnert A (1999) Interviewing as communication: An alternative way of administering the diagnostic interview schedule for children. *J Abnorm Child Psychol* 27:447–453.
- Shrout PE, et al. (2010) *Support Processes in Intimate Relationships*, eds Sullivan KT, Davila J (Oxford Univ Press, New York), pp 175–199.
- Lazarus RS, Folkman S (1984) *Stress, Appraisal, and Coping* (Springer, New York).
- Kessler RC, et al. (2002) Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychol Med* 32:959–976.
- Bird HR, et al. (2007) Longitudinal development of antisocial behaviors in young and early adolescent Puerto Rican children at two sites. *J Am Acad Child Adolesc Psychiatry* 46:5–14.
- Gunderson JG, et al. (2011) Ten-year course of borderline personality disorder: Psychopathology and function from the collaborative longitudinal personality disorders study. *Arch Gen Psychiatry* 68:827–837.
- Morey LC, Hopwood CJ (2013) Stability and change in personality disorders. *Annu Rev Clin Psychol* 9:499–528.
- Guenther PM, Dodd KW, Reedy J, Krebs-Smith SM (2006) Most Americans eat much less than recommended amounts of fruits and vegetables. *J Am Diet Assoc* 106: 1371–1379.
- Duan N, Alegria M, Canino G, McGuire TG, Takeuchi D (2007) Survey conditioning in self-reported mental health service use: Randomized comparison of alternative instrument formats. *Health Serv Res* 42:890–907.
- Gleason ME, Iida M, Shrout PE, Bolger N (2008) Receiving support as a mixed blessing: Evidence for dual effects of support on psychological outcomes. *J Pers Soc Psychol* 94: 824–838.
- McNair DM, Lorr M, Droppleman LF (1992) *EdITS Manual for the Profile of Mood States* (Educational and Industrial Testing Service, San Diego).
- Cranford JA, et al. (2006) A procedure for evaluating sensitivity to within-person change: Can mood measures in diary studies detect change reliably? *Pers Soc Psychol Bull* 32:917–929.
- Cohen J (1988) *Statistical Power Analysis for the Behavioral Sciences* (Erlbaum Associates, Hillsdale, NJ).