

Supplementary Appendix

Study	Relevant outcomes	Operationalization	Summary of the most important results
Bostock et al. 2019	Psychological Well-being and daily well-being¹	Warwick Edinburgh Mental Well-being Scale Positive emotions ratings via daily diary methodology	Intervention Effects at Primary Endpoint (t2) - Significant Group-by-time interactions: global well-being: $F(1,227)= 8.77$, $p= .003$; daily well-being: $F(1,227)= 8.37$, $p=.004$ Practice time effects (the effect of practice amount) - Significant effect of practice amount: global well-being, $F(1,121)= 3.45$, $p=.035$, daily well-being: $F(1,104)= 3.26$, $p= .042$ Intervention Effects at Second Follow-up Assessment (t3) - Consistent effect on t3: global well-being, $t(106)= -0.878$, $p= .382$
	Psychological distress	Subscales of the Hospital Anxiety and Depression Scale	Intervention Effects at Primary Endpoint (t2) - Significant Group-by-time interactions: Anxiety Symptoms $F(1,227)= 7.78$, $p= .006$; Depressive symptoms $F(1,227)= 15.6$, $p= .0001$ Practice time effects (the effect of practice amount) - Significant effect of practice amount: anxiety symptoms, $F(1,121)= 10.24$, $p= .026$; depressive symptoms, $F(1, 121)= 3.76$, $p= .026$. Intervention Effects at Second Follow-up Assessment (t3) - Consistent effect on t3: depressive symptoms, $t(105)= 0.391$, $p= .696$ - Increased effect on t3: anxiety symptoms, $t(104)= -2.053$, $p= .04$.
	Job strain	16 items extracted from the Whitehall II study questionnaire	Intervention Effects at Primary Endpoint (t2) - Significant Group-by-time interactions: Job strain $F(1,227)= 5.39$, $p= 0.21$; Job control $F(1,227)= 5.71$, $p= .018$ - Non-Significant Group-by-time interactions: Job demands $F(1,227)= 1.12$, $p= .291$ Intervention Effects at Second Follow-up Assessment (t3) - Consistent effect on t3: job strain, $t(103)= -0.152$, $p= .879$
	Workplace social support	Using five statements ranked on a 4-point scale a	Intervention Effects at Primary Endpoint (t2) - Significant Group-by-time interactions: $F(1,227)= 4.61$, $p= .003$ Practice time effects (the effect of practice amount) - No significant effect of practice amount: $p= .911$
	Blood Pressure	Healthy range (systolic BP < 120 mmHg and diastolic BP < 80 mmHg)	Intervention Effects at Primary Endpoint (t2) - No significant Group-by-time interactions of BP: Systolic BP, mmHg ^b $F(1,227)= 3.28$, $p= .071$; Diastolic BP, mmHg ^b $F(1,227)= 1.26$, $p= .0262$ Practice time effects (the effect of practice amount)

			- No significant effect of practice amount: Systolic BP, mmHg ^b $p = .261$; Diastolic BP, mmHg ^b $p = .316$
Champion et al. 2018	Distress (depression, anxiety, social impairment and hypochondriasis)	General Health Questionnaire 28 (GHQ)	Just used for Baseline measurements
	Satisfaction with life	The Satisfaction with Life Scale (SWLS)	- Significant main effect of time: $F(2,120) = 5.74$, $p = .004$ - Significant group by time interaction $F(2,120) = 5.057$, $p = .008$ of intervention group - Post-hoc paired t-tests showed that satisfaction with life increased in the MM group from baseline to day 10 ($t(28) = 2.39$, $p = .024$), and increased further from day 10 to day 30, though the latter just failed to reach significance ($t(28) = 1.84$, $p = .076$). - No significant increase of WL group across any of the assessment points (all $p > 0.05$) - Between-group effect size analysis with a Cohen's d of 0.60 (95%-CI: 0.08 to 1.12) between baseline and day 30
	Perceived Stress	The Perceived Stress Scale (PSS)	- Significant main effect of group $F(1,120) = 15.56$, $p < .001$, and an almost significant main effect of time $F(2,120) = 2.88$, $p = .06$ - Significant group by time interaction $F(2,120) = 21.98$, $p < .001$ - Post-hoc paired t-tests showed that perceived stress decreased in the MM group between baseline and day 10 ($t(28) = 2.84$, $p = .008$), and decreased further between day 10 and day 30 ($t(28) = 3.16$, $p = .004$) - No changes in stress between baseline and day 10 in the WL group ($p > .05$) but increased stress between days 10 and 30 ($t(32) = 4.49$, $p < .001$) - Between-group effect size analysis with a Cohen's d of 1.53 (95%-CI: 0.95 to 2.11) between baseline and day 30
	Resilience	The Wagnild Resilience Scale (WRS)	- Significant main effect of time $F(1,120) = 9.21$, $p < .001$ - Significant group by time interaction $F(1,120) = 5.86$, $p = .004$ - Post-hoc paired t-tests showed that resilience scores in the MM group increased between baseline and day 10 ($t(28) = 3.37$, $p = .002$), and increased further between day 10 and day 30 ($t(28) = 2.96$, $p = .006$) - Resilience scores did not differ between any assessment points in the WL group (all $p > .05$) - Between-group effect size analysis with Cohen's d of 0.61 (95%-CI: 0.09 to 1.13) between baseline and day 30
	Engagement and experience	Engagement and experience questionnaire	Just measured for the MM group - Participants in the MM group reported engaging with the app an average of 6.21 times (SD= 2.65) between baseline and day 10 (min= 1, max= 10),

			- Participants that rated the intervention as being relatively easy between days 10 and 30 were likely to have larger increases in satisfaction with life ($\beta = 1.264$, $p = .006$), larger decreases in stress ($\beta = 1.471$, $p = .015$), and larger increases in resilience ($\beta = 2.570$, $p = .003$) over the same period. - Participants with higher GHQ-28 scores at baseline (higher levels of baseline distress) were likely to have larger increases in satisfaction with life ($\beta = 0.508$, $p = .041$) and larger increases in resilience ($\beta = 1.231$, $p = .009$) between days 10 and 30 of the MM intervention. - Older participants were likely to have a larger increase in satisfaction with life between baseline and day 10 ($\beta = 0.323$, $p = 0.049$), whilst females were likely to experience larger decreases in stress between day 10 and day 30 than males ($\beta = 4.794$, $p = .019$).
Collins et al. 2019	Energetic arousal (energy, tiredness, tenseness, calmness)	Activation-Deactivation Adjective Checklist (ADACL)	Study 1: - No significant effect of condition ($F(2,42) = .29$; $p = .75$; $\eta^2 = .01$) and no overall effect of time ($F(1,42) = .22$; $p = .64$; $\eta^2 = .01$) - Significant interaction between time and condition ($F(2,42) = 3.76$; $p = .03$; $\eta^2 = .15$) -> Significant interaction is more between digital game and control conditions ($t(28) = 2.72$; $p = .01$), than between the digital game and the mindfulness conditions ($t(28) = 2.04$; $p = .05$) - No significant differences between the mindfulness and control conditions ($t(28) = 0.62$; $p = .54$) Study 2: - No significant relationship in either group ($p > 0.05$)
	Recovery (psychological detachment, relaxation, control, mastery)	Recovery experience scale	Study 1: - No significant differences between the conditions were identified ($p > 0.05$) Study 2: - The main effect of the condition was not significant ($F(1,18) = 1.93$; $p = .18$) and the main effect of the day was not significant ($F(1,18) = .00$; $p = .97$) - Significant interaction between condition and day ($F(1,18) = 9.97$; $p = .01$) - Significant interaction between condition and recovery experience in the mastery ($F(1,18) = 5.99$; $p = .02$) and relaxation subscales ($F(1,35) = 14.32$) -> Relaxation increased over time for the digital game condition and decreased for the mindfulness condition
	Enjoyment	5-item scale	Study 1: - No differences between the conditions in enjoyment ratings ($F(2,42) = 1.47$; $p = .24$) - Enjoyment was significantly, positively correlated with recovery experience ($r = .69$; $p < .001$) of Pearson correlation, including all subscales except mastery - No correlation with the enjoyment of the difference in energetic arousal between T2 and T3 ($r = .241$; $p = .11$) Study 2:

			- There were no significant differences between the conditions on either a daily level or in the average enjoyment score - Significant correlation between enjoyment and recovery experience ($r = .738$; $p = .02$), enjoyment and psychological detachment ($r = .672$; $p = .03$), mastery ($r = .676$; $p = .03$) and change in energetic arousal before and after the activity ($r = .46$; $p < .01$) in the digital game condition but not in the mindfulness condition - Significant correlation between enjoyment and the relaxation subscale ($r = .799$; $p = .01$) in the mindfulness condition but not the digital game-based condition.
Costa et al. 2018	Perception of health-related well-being	SF36v2 short form	- Significance but slight improvements in terms of general health-related well-being ($F(2.57) = 5.231$, $p = .008$) with GBLP in comparison with the CAP.
	Quality of life	WHOQOL-BREF is a 26-item scale	- Significance but slight improvements in terms of perception of mental health ($F(2.57) = 3.771$, $p = .029$) with GBLP in comparison with the CAP.
Deady et al. 2022	Depressive Symptoms	PHQ-9	- Significant group-by-time interaction for depression symptoms ($F(2,1123.5) = 3.97$, $p = .019$ at primary 3-month assessment, $F(3,734.7) = 2.98$, $p = .031$ at 12-month) - HeadGear was associated with a decline in mean PHQ-9 total score of 0.45 units more than the controls at post-intervention (Cohen's $d = 0.15$; 95%-CI 0.03–0.27) - At the 3-month follow-up, the advantage of HeadGear was slightly lower, but remained statistically significant, at 0.41 units on the PHQ-9 ($d = 0.13$; 95%-CI 0.01–0.27) - At the 12-month follow-up, group-by-time interaction for depression symptoms was not significantly different from baseline [$t(473.8) = 1.50$, $p = .135$]
	Anxiety	Generalized Anxiety Disorder scale (GAD-2)	- No Significant change between conditions from baseline to post-intervention -0.38 (95 % CI -0.28 to 0.09), $t(-1.02)$, $df = 1286.5$, $p = .321$) and to 3-month follow-up -0.09 (95 % CI -0.29 to 0.09), $t(-0.99)$, $df = 845.35$, $p = .113$ and 12-month follow-up -0.19 (95 % CI -0.43 to 0.05), $t(-0.70)$, $df = 350.00$, $p = .487$)
	Resilience	Connor–Davidson Resilience Scale (CD-RISC10)	- Significant change between conditions from baseline to 3-months follow-up (MD= 0.72 (95 % CI 0.04–1.39), $t(2.09)$, $df = 798.85$, $p = .037a$) - No significant change between conditions from baseline to post-intervention (MD= 0.30 (95 % CI -0.19 to 0.79), $t(1.20)$, $df = 1098.51$, $p = .230$) and 12-months follow-up (MD= 0.74 (95 % CI -0.22 to 1.69), $t(1.52)$, $df = 467.94$, $p = .130$)
	Well-being	WHO Wellbeing Index (WHO-5)	- Significant change between conditions from baseline to post-intervention (MD= 0.63 (95% CI 0.15–1.11), $t(2.55)$, $df = 1143.04$, $p = .011$ and to 12-months follow-up (MD= 1.13 (95 % CI -0.26–1.99), $t(2.56)$, $df = 502.70$, $p = .011$) - No Significant change between conditions from baseline to 3-month follow-up (MD= 0.63 (95 % CI -0.03 to 1.29), $t(1.85)$, $df = 822.98$, $p = .065$)

	Work performance	Three items from the Health and Work Performance Questionnaire (HPQ)	- A significant difference in mean change for work performance (HPQ) at follow-up ($t(935.4) = 3.03, p = .003$) - This was equivalent to an extra day of effective work each month amongst those using HeadGear compared to the control (1.19; 95% CI 1.10–1.29) - Significant change between conditions from baseline to 3-month follow-up (MD= 1.04 (95% CI 0.19–1.89), $t(2.41), df = 661.62, p = .016$) and from baseline to 12-month follow-up (MD= 1.73 (95% CI 0.60–2.86), $t(3.01), df = 464.10, p = .003$) - No significant change between conditions from baseline to post-intervention (MD= -0.23 (95% CI -1.32 to 0.86), $t(-0.42), df = 1166.14, p = .677$)
Economides et al. 2018	Stress (personal vulnerability and event load)	Stress Overload Scale (SOS)	- Significant main effect of time ($F(1,67) = 63.1, p < 0.001$) but no significant group $\times$ time interaction ($F(1,67) = 2.94, p = 0.09$) for the personal vulnerability - Significant effect of time ($F(1,67) = 12.33, p < 0.001$), and group $\times$ time interaction ($F(1,67) = 9.00, p = 0.004$) for event load - event load subscale scores of the SOS decreased significantly in the mindfulness group from pre- to post-intervention ($t(40) = 4.13, p < 0.001$), but did not change in the audiobook control group ($p = 0.92$). - between-group Cohen's d-effect size was 0.26 and 0.45 for the personal vulnerability and event load subscales - ITT analysis: Significant group $\times$ time interactions for the event load sub-scale ($\beta = 5.18, SE = 1.79, p = 0.004$)
	Positive and negative affect	The scale of Positive and Negative Experience (SPANE)	- Significant main effect of time ($F(1,67) = 33.1, p < 0.001$), and a group $\times$ time interaction ($F(1,67) = 6.86, p = 0.01$) for affect balance - Post hoc paired t-tests revealed that affect balance increased significantly in the mindfulness group ($t(40) = 5.59, p < 0.001$). Affect balance also increased in the audiobook control group but just failed to reach significance ($t(27) = 1.94, p = 0.06$). - The between-group Cohen's d-effect size was 0.47 - ITT analysis: Significant group $\times$ time interactions ($\beta = -4.74, SE = 1.77, p = 0.008$), no significant interaction for personal vulnerability subscale ($\beta = 2.79, SE = 1.73, p = 0.110$).
	Frustration and irritability	Brief Irritability Test (BITe)	- Significant main effect of time ($F(1,67) = 21.3, p < 0.001$), and a group $\times$ time interaction ($F(1,67) = 5.51, p = 0.022$) - Post hoc paired t-tests revealed that irritability decreased significantly in the mindfulness group from pre-post-intervention ($t(40) = 4.97, p < 0.001$), but did not change in the audiobook control group ($p = 0.25$). - The between-group Cohen's d-effect size was 0.44 - ITT analysis: Significant group $\times$ time interactions ($\beta = 2.95, SE = 1.18, p = 0.013$)
Firestone et al. 2018	Well-being	five areas of 'wellbeing' based on the individual variables	Spiritual well-being:

		that were aligned to the four pillars of the Fonofale model of health - The strongest associations were age (oldest group) ($p=0.0001$); Other Ethnicity ($p=0.0001$); being assimilated ($p=0.0001$) and marginalised ($p=0.0001$) - after adjusting for all co-variates: being of young age (25–34 years) $p=0.031$; and acculturation (assimilation and marginalised) $p=0.008$ and 0.003, respectively, sustained significantly improved relationships with this well-being factor Physical well-being: - the strongest associations were: the older age groups (35–44 ($p=0.0005$); and 45 years ($p=0.0008$); those reporting extreme obesity (BMI 40+) ($p=0.0001$), and missing obesity data ($p=0.015$); and participants who identified as being assimilated ($p=0.035$); traditional ($p=0.0005$); and marginalised ($p=0.0001$), and having a co-morbidity ($p=0.004$) - after adjusting for sex, age, ethnicity, education, cluster region and having any comorbidity, only the participants who were aged 35–44 years ($p=0.030$) retained a positive association with physical well-being. Conversely, those who rated as being 'marginalised' ($p=0.001$) had a very small negative mean difference that was significant, albeit indicating no improvement (-0.09) compared to the 'integrated' group, by the end of the trial Mental well-being: - the relationships were evident among those: in the oldest age group (45+years), $p=0.011$; participants who were from the Other Pacific Island nations ($p=0.009$) and Others ($p=0.007$); having only secondary school qualifications ($p=0.010$); and participants whose scores identified them as being assimilated ($p=0.0001$), traditional ($p=0.009$) and 'marginalised' ($p=0.0001$) - after adjusting for all variables, only acculturation sustained a significant relationship with this well-being factor. Participants that aligned with being 'assimilated' had shown a very small negative association, indicating no improvement (-0.01, 95CI: -0.51–0.49, $p=0.002$) at the end of the trial. Those participants that affiliated with being 'marginalised' had sustained a positive significant relationship and they reported a mild improvement in mental well-being (0.27, 95CI: -0.07–0.61, $p=0.004$), compared to the 'integrated' group Family/community well-being - the following positive associations with demographic factors were highlighted: from those among the oldest age group ($p=0.0001$); being from 'Other Pacific Island' nations ($p=0.002$); participants with the lowest education qualifications: 'none' and 'secondary', $p=0.021$ and $p=0.006$, respectively; those who aligned with being 'assimilated' ($p=0.004$); 'traditional' ($p=0.0007$); and 'marginalised' ($p=0.0001$), and participants from the 'rural' cluster localities ($p=0.021$) - after controlling for all co-variates, the participants with 'no education' qualifications (1.06, 95CI: 0.57–1.55, $p=0.019$) showed a large significant (positive) improvement, compared to those participants with any 'tertiary level' qualifications (0.53, 95CI: 0.11–0.95). Regarding all acculturation modes, there were significant positive
--	--	--

			improvements for the ‘marginalized’ group ($p=0.0004$), followed by the ‘assimilated’ group ($p=0.008$), compared to the ‘integrated’ group Combined well-being: - significant positive relationships with participants: in the oldest age group (45+yrs) ($p=0.0011$); being from Other Pacific Island nations ($p=0.037$), and ‘Other’ ($p=0.002$) ethnic groups; and those who rated their acculturation status as being assimilated ($p=0.0001$); traditional ($p=0.007$), and marginalized ($p=0.0001$), all significantly reported alignment with this wellbeing factor - after adjusting for all variables, significant relationships were evident among those participants that corresponded to being ‘assimilated’ (-0.10, 95CI: $-0.99-0.79$, $p=0.0001$) – showing no improvement for this well-being; and being ‘marginalized’ (0.20, 95CI: $-0.40-0.81$, $p=0.0001$), when compared to the ‘integrated’ group
Flett et al. 2019	Depressive	20-item Center for Epidemiological Studies Depression Scale (CES-D)	Time 0 to time 1: - Significant effect of Control vs. HeadSpace: -3.74 ($SE=1.02$) ($p<0.001$) and Control vs. Smiling Mind: -3.35 ($SE=1.06$) ($p<0.01$) and no significant effect of HeadSpace vs. Smiling Mind: 0.39 ($SE=1.06$) Time 0 to time 2: - Significant effect of Control vs. HeadSpace: -3.21 ($SE=1.47$) ($p<0.05$) and Control vs. Smiling Mind: -3.00 ($SE=1.48$) ($p<0.05$) and no significant effect of HeadSpace vs. Smiling Mind: 0.21 ($SE=1.45$)
	Anxiety	Hospital Anxiety and Depression Scale–Anxiety Subscale (HADS-A)	Time 0 to time 1: - No significant effect of Control vs. HeadSpace: -0.61 ($SE=.63$) and Control vs. Smiling Mind: -1.18 ($SE=.64$) and HeadSpace vs. Smiling Mind: $.14$ ($SE=.62$) Time 0 to time 2: - No significant effect of Control vs. HeadSpace: -1.05 ($SE=.43$) and Control vs. Smiling Mind: $.62$ ($SE=.44$) and HeadSpace vs. Smiling Mind: $.02$ ($SE=.44$)
	Stress	10-item Perceived Stress Scale (PSS)	Time 0 to time 1: - No significant effect of Control vs. HeadSpace: -0.48 ($SE=.66$) and Control vs. Smiling Mind: $.28$ ($SE=.69$) and HeadSpace vs. Smiling Mind: $.77$ ($SE=.69$) Time 0 to time 2: - No significant effect of Control vs. HeadSpace: -0.36 ($SE=.95$) and Control vs. Smiling Mind: $.31$ ($SE=.95$) and HeadSpace vs. Smiling Mind: $.67$ ($SE=.94$)
	Resilience	6-item Brief Resilience Scale	Time 0 to time 1: - Significant effect of Control vs. Smiling Mind: $.16$ ($SE=.08$) ($p<0.05$) and no significant effect of Control vs. HeadSpace: $.07$ ($SE=.07$) and HeadSpace vs. Smiling Mind: $.10$ ($SE=.08$) Time 0 to time 2:

			- Significant effect of Control vs. HeadSpace: .20 (SE= .10) ($p < .05$) and no significant effect of Control vs. Smiling Mind: .07 (SE= .10) and HeadSpace vs. Smiling Mind: -.13 (SE= .10)
	Flourishing (as a proxy for well-being)	8-item Flourishing Scale	Time 0 to time 1: - No significant effect of Control vs. HeadSpace: .41 (SE= .89) and Control vs. Smiling Mind: .86 (SE= .90) and HeadSpace vs. Smiling Mind: .46 (SE= .90) Time 0 to time 2: - No significant effect of Control vs. HeadSpace: .61 (SE= 1.21) and Control vs. Smiling Mind: 1.82 (SE= 1.23) and HeadSpace vs. Smiling Mind: 1.21 (SE= 1.20)
	College Adjustment	19-item College Adjustment Test	Time 0 to time 1: - Control vs. HeadSpace: Control vs. Smiling Mind: HeadSpace vs. Smiling Mind: Time 0 to time 2: - Control vs. HeadSpace: Control vs. Smiling Mind: HeadSpace vs. Smiling Mind:
	Mindfulness	12-item Cognitive Affective Mindfulness Scale-Revised	Time 0 to time 1: - Significant effect of Control vs. HeadSpace: 1.42 (SE= .57) ($p < .05$) and no significant effect of Control vs. Smiling Mind: .15 (SE= .59) and HeadSpace vs. Smiling Mind: -1.27 (SE= .60) Time 0 to time 2: - No significant effect of Control vs. HeadSpace: 1.00 (SE= .79) and Control vs. Smiling Mind: .51 (SE= .79) and HeadSpace vs. Smiling Mind: -.49 (SE= .78)
	Expectations and Perception of App Use	Rate their randomly assigned practice	Time 0 to 2: - App use during the 30-day discretionary period significantly moderated the effect of the condition on changes in depressive symptoms, anxiety, college adjustment, and mindfulness
	App Adherence	Via an online daily diary	Time 0 to 1 - High app adherence -> using their app on 8.24 days for Headspace users (SD = 2.02, range 2–10), on 8.00 days for Smiling Mind users (SD = 2.03, range 3–10) and 8.74 days for attention control app users (SD = 1.76, range 2–10), which did not differ between conditions ($F(2.205) = 2.63$, $p = .075$) Time 1 to 2: - Nearly half of all participants reported ‘never’ using their app again during that 30-day open access period (41.8% Headspace; 50.0% Smiling Mind; 53.7% Control app) ($F(2.188) = .026$, $p = .975$). Only 16.4% of Headspace, 15.4% of Smiling Mind, and 17.9% of the control app users reported using their app two or more times per week during that open access period; again, there were no differences between the conditions ($F(2.188) = .053$, $p = .948$).
Howells et al. 2016	Satisfaction with life	Satisfaction with Life Scale (SWLS)	- No statistic significant improvement ($F = .396$, $p = .530$, $\eta_p^2 = .003$)

	Subjective well-being	Flourishing scale	- No statistic significant improvement ($F = .746$, $p = .389$, $\eta_p^2 = .006$)
	Positive and negative affect	Positive and negative affect (PANAS)	- Statistic significant improvement of positive affect ($F = 9.133$, $p = .003$, $\eta_p^2 = .071$) - No statistic significant improvement of negative affect ($F = 1.246$, $p = .389$, $\eta_p^2 = .010$)
	Depression	Center for Epidemiologic Studies Depression Scale (CES-D)	- Statistic significant improvement ($F = 3.723$, $p = .05$, $\eta_p^2 = .030$)
Keemann et al. 2017²	Well-being	The Short Warwick-Edinburgh Mental Wellbeing Scale (SWEMWBS)	Study 1: - No significant main effect of group and the main effect of time on wellbeing - Significant interaction effect of game participation and time, $F(1, 58) = 4.39$, $p < 0.05$ $\eta_p^2 = 0.07$ -> experimental group: (MTime1 = 3.353 [SDTime1 = 0.67] vs. MTime2 = 3.579 [SDTime2 = 0.56]); control group: (MTime1 = 3.414 [SDTime1 = 0.49] vs. MTime2 = 3.352 [SDTime2 = 0.62]) Study 2 - No significant differences [$t(51) = -1.642$, $p = 0.11$] -> but an increase: mean wellbeing at Time 1= 3.75 (SD = 0.48); mean wellbeing at Time 2= 3.84 (SD = 0.51). - Significant increase when controlling for the degree to which employees felt the Wellbeing Game had enabled them to connect more with others [$F(1.49) = 4.212$, $p = 0.021$, $\eta_p^2 = 0.147$]
	Stress	Self-perceived stress with a single item	Study 1: - No significant main effect of game participation and the main effect of time on stress - No significant interaction effect between game participation and time point on stress Study 2: - Significant differences: [$t(51) = -2.021$, $p = 0.049$] -> mean stress score at Time 1= 2.92 (SD = 0.74), mean stress at Time 2= 2.75 (SD = 0.76)
	Stimuli Categorization Task	International Affective Picture System (IAPS)	Study 1: - Indication that those in the experimental group categorize pictures more positively in Time 2 than in Time 1 (MTime1 = -0.156 [SDTime1 = 0.167], MTime2 = -0.123 [SDTime2 = 0.164]) and for those in the control group to categorize stimuli more negatively (MTime1= -0.185 [SDTime1 = 0.179], MTime2 = -0.196 [SDTime2 = 0.181]) - No significant main effects of group or time, and no significant interaction between group and time point, ($F(1.48) = 0.33$, $p = 0.569$, $\eta_p^2 = 0.07$)
	Organizational Attitudes	Three items for measuring	Study 2: - No significant differences in any of these organizational attitudes were found.

		organizational attitudes and single-item measures were used to minimize the time commitment	
	Perceptions of Wellbeing Game	Four additional questions	No results visible
Kelders et al. 2018	Behavioural engagement	Means of usage measures gathered through system logs	Time - No significant differences between the conditions (M= 27.96 (SD= 12.64), d= -0.17, F(1,73)= 0.516, p= .48) Number of exercises - No significant differences between the conditions (M= 4.39 (SD= 3.13), d= 0.02, F(1,73)= 0.005, p= .95) Number of words - No significant differences between the conditions (M= 176.96 (SD= 107.66), d= 0.22, F(1,73)= 0.869, p= .35)
	Cognitive engagement	The short version of the Personal Involvement Inventory (PII)	- Significant differences (d= 0.55; F(1,73)= 5.919 p= .02) in real-life experiment but not pilot experiment (d=-0.53; F(1,73)=1.271; p=.28)
	Affective engagement	Positive And Negative Affect Schedule (PANAS)	- Significant higher scores of the gamified condition compared to the non-gamified version about the emotions “interest” and “inspiration,” (Z=-2.239; P=.03 and Z=-2.454; P=.01; respectively) - No significant differences in the total score of positive emotions (F(1,73)= 3.094; p= .08) and other emotions in a pilot experiment or real-life experiment - No evidence that the condition influenced involvement or flow independent of the effect on the positive emotions (involvement, c'=0.460; flow, c'=0.139)
	Enjoyment	Enjoyment subscale of the Intrinsic Motivation Inventory (IMI-E)	- No significant differences between the conditions (M=4.45 (SD= 1.33), d= 0.41, F(1,73)= 3.156 p= .08)
	To measure the both cognitive and affective engagement	Flow State Questionnaire of the Positive Psychology Lab (PPL-FSQ)	- Significant higher scores of the gamified condition compared to the non-gamified condition with approximately a medium effect size (F(1,73)= 4.626; P=.04) but not the pilot experiment outcome
	Overall satisfaction	Asking the participants to grade the	- No significant differences between the conditions (M=6.53 (SD= 1.80), d= 0.18, F(1,73)= 0.634 p= .43)

		intervention from 1 to 10	
	Usability	System Usability Scale (SUS)	- No significant differences between the conditions (M= 44.63 (SD= 18.26), d= -0.12, F(1.73)= 0.263 p=.61)
Myers et al. 2017	Multidimensional Well-being (interpersonal, community, occupational, physical, psychological, economic, overall)	21 items that define the I COPPE Scale	Multidimensional Well-being Model 0: The effect of being assigned to the FFW intervention (i.e., the intent-to-treat, ITT effect): - For each dimension of well-being, the ITT effect at T2 and at T3 was statistically non-significant and ranged from -.22 (effect size = -.14) for psychological well-being at T2 to .23 (effect size = .15) for economic wellbeing at T2. More broadly, the FFW group had an approximately equal adjusted mean for well-being as compared to the UC group at T2 and at T3 for each dimension of well-being. In summary, there was evidence that the effect of simply being assigned to the FFW intervention, without considering actual participation in the FFW intervention, was a null effect at T2 and T3 for each dimension of well-being. Model 1: The effect of being assigned to the FFW intervention for those who fully participated in the FFW intervention (i.e., the ITTc effect): Interpersonal well-being: - ITTc effect, 0.48 (effect size = .32), was not statistically significant at T2, p = .267. The ITTc effect (1.19 (effect size = .80)) was statistically significant at T3 (p = .042). The adjusted mean at T3 for compliers assigned to the UC group equalled 6.18, while the adjusted mean at T3 for compliers assigned to the FFW group equalled 7.37 Community Well-being - ITTc effect (1.16 (effect size = .71)) was statistically significant at T2 (p = .019). The adjusted mean at T2 for compliers assigned to the UC group equalled 6.78 while the adjusted mean at T2 for compliers assigned to the FFW group equalled 7.94. The ITTc effect (.94 (effect size = .59)), was statistically significant at T3 (p = .046). The adjusted mean at T3 for compliers assigned to the UC group equalled 6.94 while the adjusted mean at T3 for compliers assigned to the FFW group equalled 7.87. Occupational Well-Being - ITTc effect (.17 (effect size = .12) was not statistically significant at T2 (p = .322). The ITTc effect (.31 (effect size = .22)) was not statistically significant at T3 (p = .382). Physical Well-Being - ITTc effect (-.04 (effect size = -.03)) was not statistically significant at T2 (p = .809). The ITTc effect (.18 (effect size = .13)) was not statistically significant at T3 (p = .414). Psychological Well-Being - ITTc effect (.06 (effect size = .04)) was not statistically significant at T2 (p = .795). The ITTc effect (.81 (effect size = .56)) was statistically significant at T3 (p = .009). The adjusted mean at T3 for compliers assigned to the

			UC group equalled 5.54, while the adjusted mean at T3 for compliers assigned to the FFW group equalled 6.35. Economic Well-Being - ITTc effect (1.35 (effect size = .85)) was statistically significant at T2 ($p = .007$). The adjusted mean at T2 for compliers assigned to the UC group equalled 4.07, while the adjusted mean at T2 for compliers assigned to the FFW group equalled 5.42. The ITTc effect (1.48 (effect size = .94)) was statistically significant at T3 ($p < .001$). The adjusted mean at T3 for compliers assigned to the UC group equalled 6.12, while the adjusted mean at T3 for compliers assigned to the FFW group equalled 7.60. Overall Well-Being - ITTc effect (.00 (effect size = .00)) was not statistically significant at T2 ($p = .988$). The ITTc effect (.11 (effect size = .09)) was not statistically significant at T3 ($p = .570$) Model 2 estimated the ITTc effect and the effect of being assigned to the FFW intervention for those who did not fully participate in the FFW intervention (i.e., ITTn): - No results visible -> CACE results from model 2 were quite consistent with CACE results from model 1
	Compliance	A scoring system was created	- None of the demographic covariates were statistically significant predictors of compliance for interpersonal, occupational, physical, psychological, and overall well-being. - Earning a salary $\geq \\$50,000$, however, was a statistically significant predictor ($b = .85$, $p = .024$) of compliance for community well-being and economic well-being ($b = .84$, $p = .014$). The estimated odds of being a complier were 2.34 times (and 2.31 times) higher for higher wage earners as compared to lower wage earners for the community (and economic) well-being
Litvin et al. 2020	Resilience	The Resilience Research Centre—Adult Resilience Measure (RRC-ARM)	- Repeated-measures ANOVA of resilience: - significant main effect of intervention ($F(2.350) = 8.51$, $p < 0.001$, $\eta^2 = 0.046$) - significant interaction between intervention and time ($F(4.698) = 3.34$, $p = 0.01$, $\eta^2 = 0.019$) - not a main effect of time ($F(2.350) = 2.66$, $p = 0.07$, $\eta^2 = 0.015$) - eQuoo participants exhibited greater increases in scores compared with the CBT journal and waitlist control groups. Bonferroni-corrected post hoc comparisons indicated a significant difference in score means between the eQuoo and waitlist group ($p < 0.001$), but not between eQuoo and the CBT journal group, or the waitlist control and CBT journal group. - Within-intervention pre (t1)–post (t3) effect sizes were calculated, indicating a small effect for the eQuoo group ($M(SD)_{pre} = 49.32(5.60)$, and $M(SD)_{post} = 50.87(5.31)$, $d_{rm} = -0.37$), compared with a limited effect for the CBT journal ($M(SD)_{pre} = 48.56(6.68)$, and $M(SD)_{post} = 48.82(6.70)$, $d_{rm} = -0.06$) and waitlist groups ($M(SD)_{pre} = 47.35(7.36)$, and $M(SD)_{post} = 47.08(7.51)$, $d_{rm} = 0.06$)

			- Bonferroni-corrected post hoc tests indicated significant differences in scores between the eQuoo group and the waitlist group at t1 ($p = 0.049$), t2 ($p < 0.001$), and t3 ($p < 0.001$), whereas the difference between the CBT journal and eQuoo (or the waitlist group) was not statistically significant at any time point ($p > 0.05$)
	Psychological well-being	Ryff's Scales of Psychological Well-Being (RPRS) (PWB)	- significant main effect of intervention ($F(2.342) = 3.26$, $p = 0.04$, $\eta^2 = 0.019$) - significant interaction between intervention and time ($F(4.682) = 6.73$, $p < 0.001$, $\eta^2 = 0.038$) - not a main effect of time ($F(2.341) = 2.82$, $p = 0.06$, $\eta^2 = 0.016$) - eQuoo participants displayed greater increases in scores compared to the CBT journal and waitlist control groups. Bonferroni-corrected post hoc comparisons showed that the main effect of intervention came from the significant difference in means between the eQuoo and waitlist group ($p = 0.03$), but not between eQuoo and the CBT journal group or the waitlist control and CBT journal group. - Within-intervention pre (t1)–post (t3) effect sizes were calculated, indicating a small effect for the eQuoo group ($M(SD)_{pre} = 61.67(11.73)$, and $M(SD)_{post} = 64.53(10.74)$, $d_{rm} = -0.42$), compared to a limited effect for the CBT journal ($M(SD)_{pre} = 61.08(12.18)$, and $M(SD)_{post} = 61.15(12.71)$, $d_{rm} = -0.01$) and waitlist groups ($M(SD)_{pre} = 58.88(13.93)$, and $M(SD)_{post} = 58.35(13.111)$, $d_{rm} = 0.09$).
	Personal growth	Personal Growth Initiative Scale (PGIS)	- Significant main effect of intervention ($F(2.348) = 5.81$, $p = 0.003$, $\eta^2 = 0.03$) and time ($F(2.347) = 40.32$, $p < 0.001$, $\eta^2 = 0.19$) - Significant interaction between intervention and time ($F(4.694) = 3.99$, $p = 0.003$, $\eta^2 = 0.02$). - eQuoo participants had greater increases in scores compared with the CBT journal and waitlist control groups, and Bonferroni-corrected post hoc comparisons showed that the main effect of intervention came from the significant difference in means between the eQuoo and waitlist group ($p = 0.002$), but not between eQuoo and the CBT journal group or the waitlist control and CBT journal group - Within-intervention pre (t1)–post (t3) effect sizes were calculated, indicating a medium effect for the eQuoo group ($M(SD)_{pre} = 38.42(8.07)$, and $M(SD)_{post} = 42.15(6.43)$, $d_{rm} = -0.67$), compared with a small effect for both the CBT journal group ($M(SD)_{pre} = 37.21(8.88)$, and $M(SD)_{post} = 39.42(7.78)$, $d_{rm} = -0.48$) and the waitlist group ($M(SD)_{pre} = 36.36(7.75)$, and $M(SD)_{post} = 37.7(7.99)$, $d_{rm} = -0.25$) - Bonferroni-corrected post hoc tests indicated significant differences in scores between the eQuoo group and the waitlist group at t2 ($p = 0.02$) and t3 ($p < 0.001$), and between the CBT journal and eQuoo at t3 ($p = 0.032$).
	Anxiety	One-item anxiety scale	- ANOVA of the single-item anxiety score revealed a significant main effect of the intervention ($F(2.342) = 4.972$, $p = 0.007$, $\eta^2 = 0.03$) and time ($F(2.341) = 4.74$, $p = 0.009$, $\eta^2 = 0.03$), but not an interaction between intervention and time ($F(4.682) = 0.44$, $p = 0.78$, $\eta^2 = 0.003$) - The change in mean anxiety score across groups indicated that anxiety was highest for the waitlist group, and the change was uniform across groups.

			- The within-intervention pre (t1)–post (t3) effect sizes were small or limited for each group (eQuoo group: (M(SD) pre = 1.05(0.97), M(SD) post = 0.88(0.98), $d_{rm} = 0.20$); CBT journal: (M(SD)pre = 1.07(1.04), M(SD) post = 0.88(0.85), $d_{rm} = 0.19$); waitlist control: (M(SD)pre = 1.33(1.14), M(SD) post = 1.21(1.08), $d_{rm} = 0.12$)) - Bonferroni-corrected post hoc tests indicated significant differences in scores between the eQuoo group and the waitlist group at t1 ($p = 0.009$) and t2 ($p = 0.037$), and differences between the CBT journal and waitlist at t3 ($p = 0.038$).
	Attrition	Defined as not completing the assessments	- The odds of remaining in the study at t3 for those who completed the t1 questionnaire were 3.94 times higher in the test group compared with the control group (OR= 3.944, 95% confidence interval [CI] = 2.053–7.576, $p < 0.001$; 10% vs. 30.5% attrition) and 4.08 times higher in the test group compared with the waitlist group (OR= 4.084, 95% CI = 2.207 to 7.561, $p < 0.001$; 10% vs. 31.2% attrition).
Przybylko et al. 2021	Mental Health and Vitality	36-item Short Form Health Survey	Interaction between condition and time: - Significant improvements of the intervention group compared to the wait-list control group over time for mental health ($F(319) = 7.326$, $p = 0.007$) and vitality ($F(319) = 9.445$, $p = 0.002$) Changes from Week 1 to Week 12: - Significant improvements of the intervention group compared to the control group for mental health ($t(357) = -3.958$, $p < 0.001$, $d_z = 0.42$) and vitality ($t(357) = -4.279$, $p < 0.001$, $d_z = 0.45$) - Significant improvements within the intervention group for mental health (10%, $t(167) = -6.423$, $p < 0.001$, $d_z = 0.50$) and vitality (22%, $t(167) = -7.043$, $p > 0.001$, $d_z = 0.54$) - Significant improvements in the wait-list control group for mental health (2%, $t(190) = -2.210$, $p = 0.028$, $d_z = 0.16$) but not vitality (4%, $t(190) = -1.710$, $p = 0.089$, $d_z = 0.12$) Changes from Week 1 to Week 24: - Significant improvements of the intervention group compared to the wait-list control group for mental health ($t(319) = -2.707$, $p = 0.007$, $d_z = 0.30$) and vitality ($t(319) = -3.073$, $p = 0.002$, $d_z = 0.34$) - Significant improvements within the intervention group for mental health (8%, $t(158) = -5.375$, $p < 0.001$, $d_z = 0.43$) and vitality (19%, $t(158) = -5.997$, $p < 0.001$, $d_z = 0.48$) - No significant improvements and similar to Week 1 of the wait-list control group for mental health (2%, $t(161) = -1.382$, $p = 0.169$, $d_z = 0.11$) and vitality (5%, $t(161) = -1.357$, $p = 0.177$, $d_z = 0.11$)
	Depression, Anxiety and Stress	21-item Depression, Anxiety and Stress Scale (DASS-21)	Interaction between condition and time: - Significant improvements of the intervention group compared to the wait-list control group over time for depression ($F(319) = 7.841$, $p = 0.005$), anxiety ($F(319) = 4.440$, $p = 0.360$) and stress ($F(319) = 12.494$, $p < 0.001$) Changes from Week 1 to Week 12:

		- Significant improvements of the intervention group compared to the control group for depression ($t(357)=3.195$, $p=0.002$, $d_z=0.34$), anxiety ($t(357)=4.325$, $p<0.001$, $d_z=0.46$) and stress ($t(357)=3.508$, $p=0.001$, $d_z=0.37$) - Significant improvements within the intervention group for depression (-41%, $t(167)=6.189$, $p<0.001$, $d_z=0.48$), anxiety (-38%, $t(167)=5.030$, $p<0.001$, $d_z=0.39$) and stress (-31%, $t(167)=6.702$, $p<0.001$, $d_z=0.52$) - Significant improvements of the wait-list control group for stress (-9%, $t(190)=2.092$, $p=0.038$, $d_z=0.15$) but not depression (17%, $t(190)=1.596$, $p=0.112$, $d_z=0.12$) and anxiety (5%, $t(190)=-0.796$, $p=0.427$, $d_z=0.05$) Changes from Week 1 to Week 24: - Significant improvements of the intervention group compared to the wait-list control group for depression ($t(319)=2.800$, $p=0.005$, $d_z=0.31$), anxiety ($t(319)=2.113$, $p=0.035$, $d_z=0.24$) and stress ($t(319)=3.545$, $p<0.001$, $d_z=0.40$) - Significant improvements within the intervention group for depression (-40%, $t(158)=5.070$, $p<0.001$, $d_z=0.40$), anxiety (-29%, $t(158)=4.211$, $p<0.001$, $d_z=0.33$) and stress (-26%, $t(158)=5.808$, $p<0.001$, $d_z=0.46$) - No significant improvements and similar to Week 1 of the wait-list control group for depression (-5%, $t(161)=0.598$, $p=0.551$, $d=0.05$), anxiety (-5%, $t(161)=0.444$, $p=0.657$, $d=0.03$) and stress (1%, $t(161)=-0.112$, $p=0.991$, $d_z=0.01$)
	Satisfaction with life	5-item Satisfaction With Life Scale (SWLS) Interaction between condition and time: - Significant improvements of the intervention group compared to the wait-list control group over time ($F(319)=8.731$, $p=0.003$) Changes from Week 1 to Week 12: - Significant improvements of the intervention group compared to the control group ($t(357)=-4.702$, $p<0.001$, $d_z=0.50$) - Significant improvements within the intervention group (8%, $t(167)=-6.199$, $p<0.001$, $d_z=0.48$) - No significant improvements in the wait-list control group (0%, $t(190)=0.161$, $p=0.872$, $d_z=0.14$) Changes from Week 1 to Week 24: - Significant improvements of the intervention group compared to the wait-list control group ($t(319)=-2.958$, $p=0.003$, $d_z=0.33$) - Significant improvements within the intervention group (8%, $t(158)=-5.875$, $p<0.001$, $d_z=0.47$) - No significant improvements and similar to Week 1 of the wait-list control group (1%, $t(161)=-0.365$, $p=0.715$, $d_z=0.03$)
	Attrition	- At Week 24, the attrition rate was 24% - No significant differences between the completers and dropouts in age ($p=0.510$), sex ($p=0.511$), ethnicity ($p=0.081$) and education ($p=0.307$).

			- Dropouts had significantly poorer scores for mental health (75.4 ± 14.5 vs 59.1 ± 22.4, $t(424)=6.438$, $p<0.001$), depression (2.5 ± 2.6 vs 3.8 ± 4.1, $t(424)=-2.936$, $p=0.004$), anxiety (1.7 ± 1.9 vs 2.5 ± 2.7, $t(424)=-2.636$, $p=0.010$) and life satisfaction (23.8 ± 6.5 vs 22.2 ± 7.9, $t(424)=1.977$, $p=0.49$)
Routledge et al. 2021	Mental well-being	COMPAS-W	ITT: - significant age x group x time interactions for wellbeing ($F=4.133$; $p=.007$) but no significance in the post hoc comparisons PP: - No significant results
	Depression and anxiety	Depression, Anxiety and Stress Scale (DASS-42)	ITT: - significant age x group x time interactions for depression & anxiety symptoms ($F=3.776$; $p=.011$) but no significance in the post hoc comparisons PP: - significant age x group x time interaction for depression & anxiety symptoms ($F=3.330$; $p=.006$), but no significance in the groups in the post hoc comparisons
	Emotional cognition (Explicit emotion identification and implicit emotional bias)	WebNeuro (a series of online tasks designed to measure cognitive and emotional performance)	ITT: Implicit emotion bias: - Two significant sex x group x time interactions: one for angry minus neutral RT ($F=6.004$, $p=0.001$), and one for happy minus neutral RT ($F=5.403$, $p=0.001$) but almost for fear minus neutral RT ($F=3.780$, $p=0.011$) - One significant age x group x time interaction for sad minus neutral RT (3.745, $p=0.011$) - No significant results regarding sex x group x age and group x time interactions Explicit emotion identification: - Significant group x time interaction for training effects for fear expressions ($F=10.699$, $p<0.001$, $d=0.33$) - No significant results regarding sex x group x time x age, sex x group x time and age x group x time interactions PP: Implicit emotion bias: - Two significant sex x group x time interactions: one for angry minus neutral RT ($F=3.540$, $p=0.004$), and one for happy minus neutral RT ($F=3.233$, $p=0.007$) - No significant results regarding sex x group x age, age x group x time and group x time interactions Explicit emotion identification: - Significant sex x group x time x age interaction for fear ($F=2.317$, $p=0.002$) - Significant group x time interaction for fear expressions ($F=7.660$, $p=0.001$) - No significant results regarding sex x group x time and age x group x time interactions

Schakel et al. 2020	Self- Reported Vitality	Subjective Vitality Scale (SVS)	- No significant differences between the groups for self-reported vitality within 1 week after the intervention (measured at 3 points in time: (baseline vitality), after intervention/pre-vaccination, after vaccination, follow-up)
	Self-Reported Quality of Life	RAND-36	- Measuring 2 subscales (physical, mental); in both cases no significant group differences (P=.92 and P=.24, respectively)
	Self-Reported Bodily Sensations	Pennebaker Inventory of Limbic Languidness	- Significant main effect for the condition after the intervention/pre-vaccination (p=.042), indicating fewer bodily sensations for the intervention group than the control condition; - Significant main effect of time (p=.002); - No significant interactions between time and condition (p=.36) - Holms corrected pairwise comparisons: significant decrease from baseline to after the intervention (pre-vaccination): t(64)=3.16, adjusted p=.004; and a significant decrease from baseline to follow-up: adjusted P=.019
	Self-Reported Sleep Problems	9 items of the Medical Outcomes Study Sleep Scale	- Trend for an effect of the intervention (p=.074) -> repeated measures did not show any significance of time (p=.18); - Significant interaction between time and intervention (p=.027) - Holms: significant difference between intervention condition and control condition from baseline to after the intervention (pre-vaccination): adjusted p=.036 & from baseline to follow-up: adjusted p=.030
	Self-Reported Wellbeing	20-item Positive and Negative Affect Schedule [43] and a 7-item numeric rating scale (NRS)	Positive & negative affect: - No significant interaction between time and condition (p=.69, p=.15, respectively) Well-being: - Significant main effect of time (p<.001) + significant interaction effect between time and intervention (p=.033) - Holms comparisons: significant difference between intervention and control condition from baseline to the end of the test day: adjusted p=.024
	Psychophysiological Endpoints	Heart rate (HR), heart rate variability (HRV), and skin conductance were assessed with the BIOPAC MP150 system, saliva samples (to measure cortisol and alpha-amylase)	skin conductance: - no significant effects or interactions for time (p=.46) or time and condition (p=.26) HR: - Significant main effect of time (p<.001) - Significant decrease from baseline to the end of the test day (adjusted p<.001) - Holms corrected pairwise comparison, irrespective of conditions; the trend for an interaction effect between time and condition (p=.081, indicating lower HR at follow-up in the intervention group) HRV: - Significant main effect of time (p=.019, varied over time); no significance for the interaction effects between time and condition (p=.15) Slavia sample: - for cortisol no significant interaction between time and condition (p=.11), same for alpha-amylase (p=.90)

	Immune Endpoints	Blood samples (measurement of cytokines and chemokines, immunoglobulin G (IgG) antibody levels)	Cytokines and chemokines - for the control condition: significant increases for various cytokines and chemokines as well as significant decreases in other chemokines ($p < .05$) between pre-vaccination and end of test day - for the intervention condition: significant increases for various cytokines and chemokines from the same period ($p < .05$), as well as significant decreases for others IgG antibody levels: - no significant differences between the groups were found, but within the intervention condition there were significant increases in purified protein derivative-specific IgG levels ($p < .05$) - no significant group differences in serum c-reactive protein levels, as well as in the LPS stimulation of the whole blood sample ($P > .05$)
¹ Thick print means that the outcomes in the study are primary.			
² Only Study 1 fulfils the inclusion criteria of this study.			