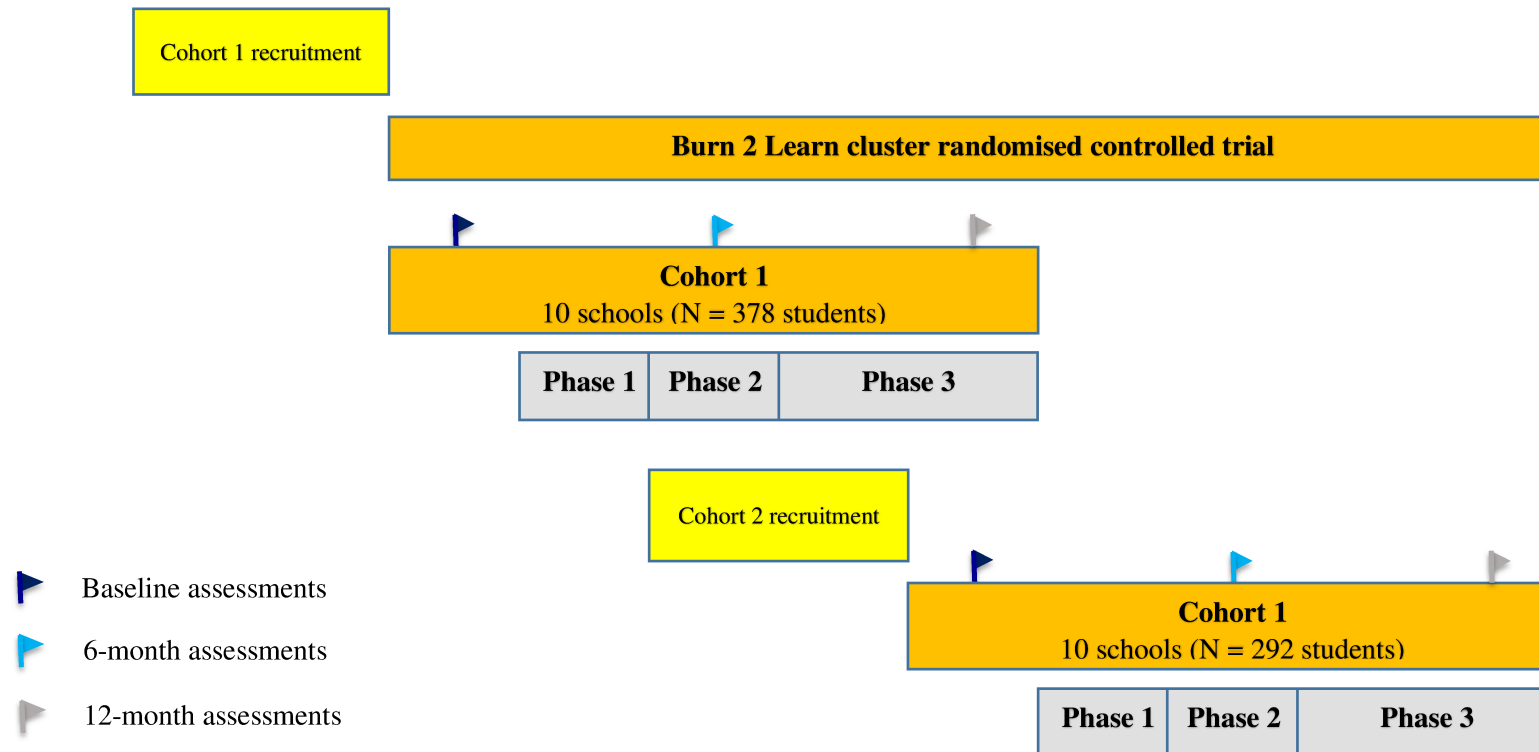


SUPPLEMENTARY MATERIAL

Supplementary Figure 1: Study timeline and assessment points

| Year | 2017 | | 2018 | | | | 2019 | | | | 2020 | |
|------|------|---|------|---|---|---|------|---|---|---|------|---|
| Term | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 |



Supplementary Table 1. Process evaluation

| | |
|---|---------------|
| 1. Intervention dose | |
| B2L sessions/week in Phase 1, mean (SD) ^a | 2.0 (0.8) |
| B2L sessions/week in Phase 2, mean (SD) ^a | 1.7 (0.6) |
| B2L sessions/week in Phase 3, mean (SD) ^a | 0.6 (0.7) |
| Total number of teacher reported B2L sessions delivered, mean (SD) ^b | 25.9 (5.2) |
| Typical length of B2L sessions, n (%) | |
| 4 minutes | 4 (19.0) |
| 8 minutes | 12 (57.1) |
| 12 minutes | 4 (19.0) |
| 16 minutes | 1 (4.8) |
| 2. Intervention fidelity (session quality)^c | |
| Adherence to SAAFE delivery principles, mean (SD) | 16.4 (2.5) |
| 3. Intervention fidelity (session intensity)^d | |
| Average HR during sessions, mean beats per minute (SD) | 143.1 (21.8) |
| Average HR during sessions, mean % of HR _{max} (SD) | 70 (11) |
| Peak HR during sessions, mean beats per minute (SD) | 167.6 (20.4) |
| Peak HR during sessions, mean % of HR _{max} (SD) | 82 (10) |
| Perceived level of exertion during B2L sessions, mean (SD) ^e | 6.3 (2.0) |
| 4. Satisfaction with the program | |
| Teacher satisfaction, mean (SD) ^f | 3.3 (0.5) |
| Student satisfaction, mean (SD) ^g | 3.8 (0.9) |
| Popularity of different B2L sessions (most to least popular) | Ranking (1-5) |
| Class HIIT | 1 |
| Quick HIIT | 2 |
| Gym HIIT | 3 |
| Sport HIIT | 4 |
| Custom HIIT | 5 |
| 5. Sustainability | |
| Participation in future HIIT (student), Yes, % | 69.6 |
| Delivery of B2L program to future student cohorts (teacher), Yes, % | 81.8 |

Note. B2L = Burn 2 Learn; SAAFE = Supportive, Active, Autonomous, Fair and Enjoyable; HR = heart rate; SD = standard deviation; HIIT = high-intensity interval training

^a Teachers asked to retrospectively report number of sessions they delivered in Phases 1-3.

^b Number of sessions delivered in Phases 1-2, logged by teachers in B2L handbook.

^c Observations of session quality of scored on a 4-point Likert scale- Strongly disagree (1) to Strongly agree (4), total /20.

^d Mean peak heart rate (% HR_{max}) and mean heart rate for the entire session (% HR_{max}) extracted from the B2L app.

^e Scored on an 11-point Likert scale ranging from Extremely easy (0) to Extremely hard (10)

^f Scored on a 4-point Likert scale ranging from Strongly disagree (1) to Strongly agree (4)

^g Scored on a 5-point Likert scale ranging from Poor (1) to Excellent (5)

Supplementary Table 2. Baseline, 6- and 12-month statistics for fitness and physical activity outcomes

| Variable | Statistic | Baseline | | 6-months | | 12-months | |
|----------------------------------|-------------------|-----------------------|-----------------------|----------------------|-----------------------|----------------------|-----------------------|
| | | CON | INT | CON | INT | CON | INT |
| Cardiorespiratory fitness (laps) | mean (SD) | 49.9 (27.3) | 47.8 (24.5) | 50.3 (29.1) | 51.0 (24.0) | 46.5 (27.5) | 45.5 (21.2) |
| | median (min, max) | 49.0 (8.0, 123.0) | 43.0 (6.0, 109.0) | 46.0 (6.0, 138.0) | 49.0 (7.0, 116.0) | 42.0 (5.0, 145.0) | 42.0 (8.0, 98.0) |
| Upper body endurance (reps) | mean (SD) | 11.7 (9.6) | 11.1 (8.4) | 12.2 (9.9) | 12.2 (8.8) | 12.7 (10.2) | 12.9 (8.8) |
| | median (min, max) | 10.0 (0.0, 41.0) | 10.0 (0.0, 36.0) | 10.0 (0.0, 40.0) | 11.0 (0.0, 37.0) | 10.0 (0.0, 41.0) | 13.0 (1.0, 35.0) |
| Lower body power (cm) | mean (SD) | 172.3 (39.5) | 174.1 (36.8) | 174.3 (39.0) | 174.9 (36.8) | 180.1 (43.0) | 170.8 (39.1) |
| | median (min, max) | 168.5 (84.0, 281.0) | 174.0 (90.0, 260.0) | 176.0 (95.0, 274.0) | 170.5 (99.0, 260.0) | 179.0 (73.0, 293.0) | 167.0 (66.0, 260.0) |
| BMI z-score | mean (SD) | 0.78 (1.11) | 0.75 (0.98) | 0.76 (1.08) | 0.82 (0.92) | 0.72 (1.13) | 0.76 (0.96) |
| | median (min, max) | 0.68 (-2.05, 3.74) | 0.74 (-1.32, 3.22) | 0.77 (-2.65, 3.70) | 0.80 (-1.28, 3.23) | 0.66 (-2.50, 3.65) | 0.78 (-1.13, 3.14) |
| MPA mins/school hours | mean (SD) | 17.6 (7.2) | 19.0 (8.6) | 15.6 (8.6) | 17.7 (9.3) | 15.3 (7.8) | 14.5 (9.6) |
| | median (min, max) | 16.7 (3.6, 42.2) | 17.6 (2.2, 54.2) | 13.9 (0.3, 48.6) | 16.4 (0.3, 61.6) | 13.2 (0.2, 45.5) | 13.1 (0.0, 40.6) |
| VPA mins/school hours | mean (SD) | 0.7 (0.8) | 0.8 (1.2) | 0.8 (1.1) | 0.9 (1.3) | 0.6 (1.1) | 0.7 (1.2) |
| | median (min, max) | 0.5 (0.0, 5.9) | 0.4 (0.0, 9.1) | 0.3 (0.0, 6.6) | 0.5 (0.0, 9.3) | 0.3 (0.0, 8.5) | 0.2 (0.0, 6.6) |
| MVPA mins/school hours | mean (SD) | 18.4 (7.6) | 19.9 (9.3) | 16.3 (9.4) | 18.6 (10.1) | 15.9 (8.3) | 15.9 (9.7) |
| | median (min, max) | 17.1 (3.6, 42.7) | 18.3 (2.6, 60.2) | 14.3 (0.3, 54.4) | 17.0 (0.3, 69.1) | 13.6 (0.2, 47.8) | 13.7 (0.0, 44.5) |
| Steps/school hours | mean (SD) | 6,560 (2,219) | 6,157 (1,506) | 5,562 (1,388) | 5,923 (1,903) | 5,611 (1,381) | 5,222 (1,802) |
| | median (min, max) | 6,108 (3,573, 17,810) | 5,911 (1,668, 11,507) | 5,575 (1,513, 9,113) | 5,489 (1,692, 12,482) | 5,609 (1,042, 9,417) | 5,311 (1,354, 10,092) |
| MPA mins/weekday | mean (SD) | 36.1 (13.6) | 36.9 (13.7) | 32.7 (13.4) | 33.2 (12.7) | 31.6 (11.1) | 32.8 (15.3) |
| | median (min, max) | 34.5 (5.4, 81.8) | 35.0 (3.5, 83.2) | 32.1 (0.3, 84.0) | 32.5 (2.3, 76.4) | 29.0 (3.6, 61.1) | 31.2 (0.2, 88.5) |

| Variable | Statistic | Baseline | | 6-months | | 12-months | |
|-----------------------|-------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
| | | CON | INT | CON | INT | CON | INT |
| VPA mins/weekday | mean (SD) | 1.6 (1.7) | 1.7 (2.1) | 1.6 (1.8) | 1.5 (1.9) | 1.4 (1.5) | 1.2 (1.7) |
| | median (min, max) | 1.1 (0.0, 12.4) | 1.0 (0.0, 14.7) | 0.9 (0.0, 8.0) | 1.0 (0.0, 12.9) | 0.8 (0.0, 7.9) | 0.6 (0.0, 7.6) |
| MVPA mins/weekday | mean (SD) | 37.7 (14.2) | 38.6 (14.8) | 34.4 (14.4) | 34.8 (13.8) | 33.0 (11.3) | 34.1 (16.2) |
| | median (min, max) | 35.8 (5.4, 82.8) | 36.5 (4.3, 97.9) | 32.8 (0.3, 91.1) | 34.0 (2.3, 84.4) | 31.2 (3.9, 61.3) | 32.1 (0.2, 94.5) |
| Steps/weekday | mean (SD) | 11,392 (2,050) | 10,858 (2,107) | 10,919 (2,171) | 10,479 (2,229) | 10,620 (1,972) | 10,016 (2,572) |
| | median (min, max) | 11,240 (6,502, 18,085) | 10,672 (2,385, 17,263) | 10,775 (4,959, 16,658) | 10,398 (3,928, 16,859) | 10,192 (5,598, 16,951) | 10,067 (2,091, 14,640) |
| MPA mins/weekend day | mean (SD) | 33.5 (19.8) | 33.9 (17.0) | 29.6 (18.5) | 33.0 (18.7) | 27.9 (17.0) | 34.6 (17.9) |
| | median (min, max) | 29.6 (5.9, 135.1) | 31.5 (7.0, 118.8) | 25.7 (5.1, 105.4) | 29.1 (4.0, 102.6) | 22.1 (0.7, 89.6) | 30.4 (6.7, 101.0) |
| VPA mins/weekend day | mean (SD) | 0.6 (1.0) | 0.8 (1.3) | 0.7 (1.3) | 0.7 (1.5) | 0.8 (1.8) | 0.4 (0.6) |
| | median (min, max) | 0.2 (0.0, 6.1) | 0.3 (0.0, 6.5) | 0.2 (0.0, 7.7) | 0.2 (0.0, 10.4) | 0.2 (0.0, 12.2) | 0.2 (0.0, 2.8) |
| MVPA mins/weekend day | mean (SD) | 34.1 (20.2) | 34.7 (17.7) | 30.4 (18.9) | 33.7 (19.5) | 28.7 (17.3) | 35.0 (17.9) |
| | median (min, max) | 30.5 (5.9, 138.8) | 31.9 (7.0, 123.2) | 25.8 (5.1, 105.9) | 29.5 (4.0, 113.0) | 23.0 (0.7, 89.8) | 30.5 (6.8, 101.0) |
| Steps/weekend day | mean (SD) | 9,403 (2,999) | 9,623 (3,021) | 9,418 (3,577) | 9,857 (3,723) | 8,505 (3,311) | 9,925 (3,186) |
| | median (min, max) | 9,331 (4,206, 19,673) | 9,629 (3,212, 17,519) | 9,173 (2,139, 20,427) | 9,000 (3,913, 25,292) | 8,555 (2,533, 19,085) | 9,474 (4,388, 16,661) |

Note. CON = control; INT = intervention; BMI z-score = body mass index z-score; MPA = moderate physical activity; VPA = vigorous physical activity; MVPA = moderate-to-vigorous physical activity.

Supplementary Table 3. Baseline, 6- and 12-month statistics for mental health outcomes

| Variable | Statistic | Baseline | | 6-months | | 12-months | |
|-------------------------------------|-------------------|------------------|------------------|------------------|-------------------|------------------|------------------|
| | | CON | INT | CON | INT | CON | INT |
| Hair cortisol concentration (pg/mg) | mean (SD) | 10.6 (6.2) | 14.6 (12.0) | 13.0 (5.6) | 13.3 (5.3) | - | - |
| | median (min, max) | 8.9 (2.5, 37.4) | 10.8 (1.7, 68.5) | 11.6 (4.6, 33.0) | 12.4 (0.0, 35.4) | | |
| Perceived stress | mean (SD) | 18.8 (6.2) | 19.1 (6.1) | 18.1 (6.5) | 18.4 (5.8) | 19.6 (6.2) | 19.5 (6.4) |
| | median (min, max) | 19.0 (1.0, 40.0) | 19.0 (2.0, 37.0) | 18.0 (1.0, 34.0) | 19.0 (1.0, 40.0) | 20.0 (0.0, 38.0) | 19.0 (2.0, 40.0) |
| Internalising problems | mean (SD) | 5.5 (3.2) | 5.4 (3.2) | 5.6 (3.4) | 5.3 (3.1) | 5.6 (3.3) | 5.3 (3.1) |
| | median (min, max) | 5.0 (0.0, 18.0) | 5.0 (0.0, 17.0) | 5.0 (0.0, 18.0) | 5.0 (0.0, 14.0) | 5.0 (0.0, 15.0) | 5.0 (0.0, 16.0) |
| Externalising problems | mean (SD) | 6.1 (3.3) | 6.1 (3.4) | 6.1 (3.4) | 6.1 (3.2) | 6.0 (3.2) | 5.8 (3.2) |
| | median (min, max) | 6.0 (0.0, 16.0) | 6.0 (0.0, 17.0) | 6.0 (0.0, 18.0) | 6.0 (0.0, 16.0) | 6.0 (0.0, 16.0) | 6.0 (0.0, 15.0) |
| Well-being | mean (SD) | 23.9 (5.0) | 24.9 (4.8) | 24.5 (5.0) | 24.8 (4.5) | 24.6 (4.9) | 25.0 (4.5) |
| | median (min, max) | 25.0 (7.0, 35.0) | 25.0 (9.0, 35.0) | 25.0 (8.0, 35.0) | 24.0 (14.0, 35.0) | 24.0 (7.0, 35.0) | 25.0 (9.0, 35.0) |
| HIIT self-efficacy | mean (SD) | 6.2 (2.1) | 6.3 (2.2) | 6.3 (2.2) | 7.2 (2.2) | 6.3 (2.2) | 7.2 (2.3) |
| | median (min, max) | 6.4 (1.0, 10.0) | 6.4 (1.0, 10.0) | 6.5 (1.0, 10.0) | 7.5 (1.0, 10.0) | 6.5 (1.0, 10.0) | 7.3 (1.0, 10.0) |
| Intrinsic motivation for exercise | mean (SD) | 2.9 (1.0) | 2.9 (1.0) | 2.9 (1.1) | 2.9 (0.9) | 2.9 (1.0) | 2.9 (1.0) |
| | median (min, max) | 3.0 (0.0, 4.0) | 3.0 (0.0, 4.0) | 3.0 (0.0, 4.0) | 3.0 (0.0, 4.0) | 3.0 (0.0, 4.0) | 3.0 (0.0, 4.0) |
| Identified motivation for exercise | mean (SD) | 2.9 (0.9) | 2.9 (0.9) | 2.9 (0.9) | 2.9 (0.9) | 3.0 (0.9) | 3.0 (0.9) |
| | median (min, max) | 3.0 (0.0, 4.0) | 3.0 (0.0, 4.0) | 3.0 (0.0, 4.0) | 3.0 (0.0, 4.0) | 3.3 (0.0, 4.0) | 3.3 (0.0, 4.0) |

Note. CON = control; INT = intervention; pg/mg = picograms of cortisol per milligram of hair; HIIT = high-intensity interval training.

Supplementary Table 4. Baseline, 6- and 12-month means and standard deviation for cognitive outcomes

| Variable | Statistic | Baseline | | 6-months | | 12-months | |
|----------------------------|-------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| | | CON | INT | CON | INT | CON | INT |
| Congruent, accuracy (%) | mean (SD) | 92.7 (7.8) | 93.1 (7.7) | 96.5 (5.6) | 97.6 (3.2) | 97.1 (5.3) | 98.0 (3.1) |
| | median (min, max) | 96.0 (56.0, 100.0) | 96.0 (52.0, 100.0) | 98.7 (66.7, 100.0) | 98.7 (82.7, 100.0) | 98.7 (61.3, 100.0) | 98.7 (72.0, 100.0) |
| Incongruent, accuracy (%) | mean (SD) | 76.9 (14.4) | 79.0 (12.7) | 85.7 (12.0) | 87.9 (10.2) | 88.4 (11.4) | 90.3 (8.8) |
| | median (min, max) | 80.0 (10.7, 98.7) | 81.3 (26.7, 98.7) | 89.3 (25.3, 100.0) | 90.7 (37.3, 100.0) | 92.0 (42.7, 100.0) | 92.0 (40.0, 100.0) |
| Interference, accuracy (%) | mean (SD) | 15.8 (11.3) | 14.1 (9.1) | 10.8 (9.1) | 9.7 (8.7) | 8.7 (8.6) | 7.7 (7.5) |
| | median (min, max) | 13.3 (-6.7, 82.7) | 12.0 (-2.7, 57.3) | 8.0 (-4.0, 54.7) | 6.7 (-8.0, 48.0) | 6.7 (-6.7, 50.7) | 6.0 (-2.7, 45.3) |
| Congruent, RT (ms) | mean (SD) | 419.5 (53.4) | 415.9 (50.9) | 420.0 (48.4) | 421.7 (46.5) | 419.8 (47.1) | 418.2 (41.1) |
| | median (min, max) | 409.2 (298.4, 637.7) | 407.9 (316.3, 605.6) | 413.5 (312.8, 602.8) | 416.3 (333.7, 611.8) | 416.7 (303.0, 609.1) | 414.0 (332.6, 555.3) |
| Incongruent, RT (ms) | mean (SD) | 481.4 (57.8) | 477.3 (56.7) | 480.2 (55.0) | 479.9 (48.4) | 474.5 (47.5) | 470.6 (41.2) |
| | median (min, max) | 477.5 (315.7, 680.7) | 476.0 (329.5, 720.6) | 473.2 (350.9, 757.1) | 475.3 (372.0, 737.7) | 468.3 (343.9, 638.4) | 466.3 (373.3, 584.6) |
| Interference, RT (ms) | mean (SD) | 61.9 (25.9) | 61.4 (27.9) | 60.2 (27.2) | 58.2 (22.9) | 54.7 (23.3) | 52.4 (20.0) |
| | median (min, max) | 60.4 (-35.5, 141.1) | 60.2 (-19.3, 147.9) | 58.7 (-9.2, 259.9) | 56.9 (-7.3, 193.4) | 53.8 (-17.8, 168.5) | 50.4 (-5.4, 108.3) |
| Non-target, accuracy (%) | mean (SD) | 84.3 (14.4) | 85.0 (12.7) | 89.9 (10.3) | 89.3 (12.1) | 90.0 (12.3) | 90.5 (10.8) |
| | median (min, max) | 89.1 (18.5, 100.0) | 89.1 (25.0, 100.0) | 93.5 (31.5, 100.0) | 92.4 (0.0, 100.0) | 94.6 (28.3, 100.0) | 94.6 (37.0, 100.0) |
| Target, accuracy (%) | mean (SD) | 74.8 (16.7) | 75.4 (14.5) | 81.2 (13.8) | 82.6 (14.9) | 83.3 (15.6) | 84.3 (13.3) |
| | median (min, max) | 79.2 (14.6, 100.0) | 79.2 (12.5, 100.0) | 85.4 (35.4, 100.0) | 87.5 (0.0, 100.0) | 87.5 (27.1, 100.0) | 87.5 (33.3, 100.0) |
| d-prime | mean (SD) | 2.0 (1.0) | 2.0 (0.8) | 1.8 (1.0) | 1.9 (1.0) | 2.7 (1.1) | 2.7 (1.0) |
| | median (min, max) | 2.1 (-0.8, 4.5) | 2.0 (-0.2, 4.2) | 1.9 (-1.3, 3.9) | 2.0 (-1.1, 3.9) | 2.8 (-1.1, 4.6) | 2.8 (0.0, 4.5) |

| Variable | Statistic | Baseline | | 6-months | | 12-months | |
|---------------------|-------------------|-----------------------|-----------------------|-----------------------|---------------------|-----------------------|-----------------------|
| | | CON | INT | CON | INT | CON | INT |
| Non-target, RT (ms) | mean (SD) | 830.5 (209.1) | 828.2 (201.0) | 772.5 (183.3) | 767.4 (178.8) | 759.3 (191.2) | 735.4 (172.7) |
| | median (min, max) | 831.8 (288.5, 1416.2) | 825.2 (292.5, 1521.8) | 764.3 (337.5, 1376.4) | 773.6 (0.0, 1208.1) | 738.8 (334.9, 1359.1) | 727.5 (331.8, 1532.3) |
| Target, RT (ms) | mean (SD) | 723.6 (204.1) | 707.2 (179.8) | 673.4 (164.8) | 665.4 (156.1) | 673.8 (178.0) | 646.3 (164.7) |
| | median (min, max) | 719.3 (263.5, 1372.9) | 706.2 (264.8, 1248.5) | 657.9 (304.4, 1132.9) | 658.7 (0.0, 1143.0) | 658.7 (282.8, 1596.8) | 617.9 (295.5, 1379.3) |

Note. RT = reaction time; ms = milliseconds. Consistent with previous research, participants were included in the analysis if their overall mean flanker accuracy (i.e., congruent and incongruent) was higher than 50% and if their mean d-prime score was greater than 0 for the 1- and 2-back conditions. At baseline (37 cases), 6-month (13 cases) and 12-months (11 cases) flanker accuracy and reaction values were removed. At baseline (8 cases), 6-month (3 cases) and 12-months (2 cases) n-back accuracy and reaction values were removed.

Supplementary Table 5. Effect modification of CRF at 6-months

| Moderator | Interaction P value | Subgroup | Within group mean change (95% CI) (6-month minus baseline) | | Adjusted difference ¹ (INT v. CON) |
|----------------------|------------------------|--------------------------------------|---|-------------------|--|
| | | | CON | INT | Estimate |
| SES | 0.221 | High | -3.0 (-5.9 to -0.0) | 3.5 (0.4 to 6.6) | 6.5 (2.2 to 10.7) |
| | | Low or medium | -3.2 (-5.3 to -1.1) | 0.1 (-1.8 to 2.0) | 3.3 (0.5 to 6.1) |
| Sex | 0.986 | Boy | -3.4 (-5.6 to -1.6) | 0.4 (-1.9 to 2.6) | 4.0 (1.0 to 7.0) |
| | | Girl | -2.3 (-5.3 to 0.8) | 1.7 (-0.5 to 4.0) | 4.0 (0.2 to 7.8) |
| Weight status | 0.252 | Healthy weight and underweight | -2.6 (-4.6 to -0.6) | 0.7 (-1.2 to 2.3) | 3.3 (0.6 to 6.0) |
| | | Overweight and obese | -4.4 (-7.5 to -1.4) | 1.9 (-1.3 to 5.1) | 6.4 (1.9 to 10.8) |
| Mental health status | 0.483 | Close to average and slightly raised | -3.3 (-5.1 to -1.5) | 0.4 (-1.3 to 2.2) | 3.7 (1.2 to 6.3) |
| | | High to very high | -1.9 (-6.0 to 2.2) | 4.1 (0.0 to 8.2) | 6.0 (0.2 to 11.7) |
| CRF status | 0.478 | Health risk and needs improvement | -0.5 (-3.2 to 2.2) | 2.5 (-0.3 to 5.3) | 3.0 (-0.9 to 6.9) |
| | | Healthy fitness zone | -4.5 (-6.6 to -2.5) | 0.2 (-1.7 to 2.1) | 4.7 (1.9 to 7.6) |

Note. INT = intervention; CON = control; SES = socio-economic status based on household postcode; CRF = cardiorespiratory fitness; CI = confidence interval

¹ Adjusted difference at 6-months = [(INT post-test mean minus INT baseline mean) minus (CON post-test mean minus CON baseline mean)].

Supplementary Table 6. Effect modification of hair cortisol concentrations at 6-months

| Moderator | Interaction P value | Subgroup | Within group mean change (95% CI) (6-month minus baseline) | | Adjusted difference ¹ (INT v. CON) |
|----------------------|------------------------|--------------------------------------|---|------------------------|--|
| | | | CON | INT | Estimate |
| SES | 0.139 | High | 5.20 (-0.05 to 10.45) | -3.13 (-7.80 to 1.53) | -8.34 (-15.36 to -1.32) |
| | | Low or medium | 2.00 (-0.22 to 4.21) | -0.58 (-2.65 to 1.50) | -2.57 (-5.61 to 0.46) |
| Sex | 0.467 | Boy | 2.45 (-0.33 to 5.22) | -0.37 (-3.14 to 2.40) | -2.82 (-6.74 to 1.10) |
| | | Girl | 2.62 (-0.68 to 5.92) | -2.35 (-5.12 to 0.41) | -4.97 (-9.28 to -0.67) |
| Weight status | 0.049 | Healthy weight and underweight | 2.12 (-0.39 to 4.63) | 0.02 (-2.23 to 2.27) | -2.10 (-5.47 to 1.26) |
| | | Overweight and obese | 3.44 (-0.46 to 7.34) | -5.13 (-8.97 to -1.30) | -8.57 (-14.05 to -3.10) |
| Mental health status | 0.087 | Close to average and slightly raised | 2.15 (-0.16 to 4.47) | -0.60 (-2.75 to 1.56) | -2.75 (-5.91 to 0.41) |
| | | High to very high | 4.01 (-1.11 to 9.13) | -5.34 (-9.92 to -0.76) | -9.35 (-16.22 to -2.47) |
| CRF status | 0.429 | Health risk and needs improvement | 1.35 (-2.75 to 5.45) | -4.68 (-8.61 to -0.75) | -6.03 (-11.71 to -0.35) |
| | | Healthy fitness zone | 2.83 (0.22 to 5.44) | -0.53 (-2.83 to 1.77) | -3.36 (-6.83 to 0.12) |

Note. INT = intervention; CON = control; SES = socio-economic status based on household postcode; CRF = cardiorespiratory fitness; CI = confidence interval

¹ Adjusted difference at 6-months = [(INT post-test mean minus INT baseline mean) minus (CON post-test mean minus CON baseline mean)].

Supplementary Table 7. Effect modification of internalising problems at 6-months

| Moderator | Interaction P value | Subgroup | Within group mean change (95% CI) (6-month minus baseline) | | Adjusted difference ¹ (INT v. CON) |
|----------------------|------------------------|--------------------------------------|---|------------------------|--|
| | | | CON | INT | Estimate |
| SES | 0.793 | High | 0.14 (-0.35 to 0.63) | 0.09 (-0.44 to 0.61) | -0.05 (-0.77 to 0.67) |
| | | Low or medium | 0.15 (-0.21 to 0.51) | -0.02 (-0.36 to 0.32) | -0.17 (-0.66 to 0.32) |
| Sex | 0.501 | Boy | 0.19 (-0.18 to 0.55) | -0.10 (-0.50 to 0.31) | -0.28 (-0.83 to 0.26) |
| | | Girl | 0.10 (-0.37 to 0.57) | 0.09 (-0.30 to 0.49) | 0.00 (-0.61 to 0.61) |
| Weight status | 0.010 | Healthy weight and underweight | -0.12 (-0.47 to 0.23) | 0.10 (-0.22 to 0.42) | 0.22 (-0.26 to 0.69) |
| | | Overweight and obese | 0.70 (0.18 to 1.22) | -0.28 (-0.85 to 0.30) | -0.97 (-1.75 to -0.20) |
| Mental health status | 0.156 | Close to average and slightly raised | 0.42 (0.12 to 0.73) | 0.35 (0.06 to 0.64) | -0.07 (-0.49 to 0.35) |
| | | High to very high | -0.96 (-1.57 to -0.36) | -1.76 (-2.44 to -1.08) | -0.80 (-1.70 to 0.11) |
| CRF status | 0.853 | Health risk and needs improvement | 0.19 (-0.31 to 0.69) | 0.01 (-0.50 to 0.51) | -0.19 (-0.90 to 0.52) |
| | | Healthy fitness zone | 0.23 (-0.17 to 0.62) | -0.05 (-0.41 to 0.32) | -0.27 (-0.81 to 0.26) |

Note. INT = intervention; CON = control; SES = socio-economic status based on household postcode; CRF = cardiorespiratory fitness; CI = confidence interval

¹ Adjusted difference at 6-months = [(INT post-test mean minus INT baseline mean) minus (CON post-test mean minus CON baseline mean)].

Supplementary Table 8. Effect modification of perceived stress at 6-months

| Moderator | Interaction P value | Subgroup | Within group mean change (95% CI) (6-month minus baseline) | | Adjusted difference ¹ (INT v. CON) |
|----------------------|------------------------|--------------------------------------|---|------------------------|--|
| | | | CON | INT | Estimate |
| SES | 0.241 | High | -0.27 (-1.34 to 0.79) | -1.10 (-2.26 to 0.05) | -0.83 (-2.40 to 0.74) |
| | | Low or medium | -0.61 (-1.40 to 0.17) | -0.31 (-1.04 to 0.42) | 0.30 (-0.77 to 1.37) |
| Sex | 0.425 | Boy | -0.41 (-1.21 to 0.39) | -0.82 (-1.70 to 0.06) | -0.41 (-1.60 to 0.78) |
| | | Girl | -0.59 (-1.61 to 0.43) | -0.27 (-1.13 to 0.59) | 0.32 (-1.01 to 1.65) |
| Weight status | 0.032 | Healthy weight and underweight | -0.69 (-1.45 to 0.07) | -0.10 (-0.81 to 0.60) | 0.59 (-0.44 to 1.63) |
| | | Overweight and obese | -0.17 (-1.30 to 0.97) | -1.74 (-2.99 to -0.49) | -1.57 (-3.26 to 0.12) |
| Mental health status | 0.077 | Close to average and slightly raised | -0.66 (-1.37 to 0.04) | -0.30 (-0.96 to 0.37) | 0.37 (-0.60 to 1.34) |
| | | High to very high | 0.16 (-1.23 to 1.55) | -1.54 (-3.09 to 0.02) | -1.69 (-3.78 to 0.39) |
| CRF status | 0.271 | Health risk and needs improvement | 0.15 (-0.93 to 1.22) | -0.46 (-1.55 to 0.64) | -0.60 (-2.14 to 0.93) |
| | | Healthy fitness zone | -1.02 (-1.87 to -0.17) | -0.54 (-1.32 to 0.23) | 0.48 (-0.67 to 1.62) |

Note. INT = intervention; CON = control; SES = socio-economic status based on household postcode; CRF = cardiorespiratory fitness; CI = confidence interval

¹ Adjusted difference at 6-months [(INT post-test mean minus INT baseline mean) minus (CON post-test mean minus CON baseline mean)].

Supplementary Table 9. Changes in cognitive outcomes at 6- and 12-month follow-up between participants randomised to control or the B2L intervention

| Secondary outcomes | No of clusters (participants) | | Mean change from baseline (95% CI) | | Adjusted difference in follow-up ¹ | |
|---|-------------------------------|----------|------------------------------------|------------------------|---|---------|
| | CON | INT | CON | INT | Coefficient (95% CI) | P value |
| Flanker congruent, accuracy (%): 6-months | 10 (320) | 10 (326) | 4.20 (3.31, 5.09) | 4.29 (3.49, 5.09) | 0.45 (-0.72, 1.63) | 0.449 |
| Flanker congruent, accuracy (%): 12-months | 10 (327) | 10 (328) | 3.97 (2.94, 5.00) | 4.79 (3.79, 5.80) | 0.66 (-0.52, 1.85) | 0.273 |
| Flanker incongruent, accuracy (%): 6-months | 10 (320) | 10 (326) | 9.68 (8.25, 11.11) | 8.75 (7.58, 9.93) | -0.54 (-2.35, 1.28) | 0.563 |
| Flanker incongruent, accuracy (%): 12-months | 10 (327) | 10 (328) | 11.81 (10.06, 13.56) | 11.51 (10.03, 12.98) | 0.04 (-1.91, 1.99) | 0.967 |
| Flanker interference, accuracy (%): 6-months | 10 (320) | 10 (326) | -5.48 (-6.77, -4.18) | -4.47 (-5.50, -3.44) | 0.81 (-0.80, 2.41) | 0.324 |
| Flanker interference, accuracy (%): 12-months | 10 (327) | 10 (328) | -7.84 (-9.32, -6.35) | -6.71 (-7.87, -5.56) | 0.62 (-1.05, 2.30) | 0.466 |
| Flanker congruent, RT (%): 6-months | 10 (320) | 10 (326) | 1.82 (-3.12, 6.76) | 7.67 (2.88, 12.46) | 5.91 (-0.82, 12.64) | 0.086 |
| Flanker congruent, RT (%): 12-months | 10 (327) | 10 (328) | 5.09 (-0.64, 10.82) | 5.90 (0.55, 11.25) | 0.22 (-6.93, 7.38) | 0.951 |
| Flanker incongruent, RT (%): 6-months | 10 (320) | 10 (326) | 1.20 (-4.98, 7.37) | 2.70 (-2.43, 7.83) | 2.70 (-5.07, 10.47) | 0.496 |
| Flanker incongruent, RT (%): 12-months | 10 (327) | 10 (328) | -4.42 (-11.05, 2.22) | -4.90 (-10.55, 0.75) | -2.41 (-10.52, 5.69) | 0.560 |
| Flanker interference, RT (%): 6-months | 10 (320) | 10 (326) | -0.62 (-4.64, 3.40) | -4.97 (-7.95, -1.99) | -2.48 (-7.25, 2.29) | 0.308 |
| Flanker interference, RT (%): 12 months | 10 (327) | 10 (328) | -9.51 (-13.16, -5.86) | -10.80 (-14.26, -7.35) | -2.36 (-7.20, 2.47) | 0.339 |
| 2-back non-target, accuracy (%): 6-months | 10 (328) | 10 (337) | 5.03 (3.46, 6.60) | 4.24 (2.87, 5.61) | -1.08 (-3.12, 0.96) | 0.298 |
| 2-back non-target, accuracy (%): 12-months | 10 (332) | 10 (337) | 4.83 (2.91, 6.75) | 4.56 (2.72, 6.41) | -0.27 (-2.69, 2.15) | 0.825 |
| 2-back target, accuracy (%): 6-months | 10 (328) | 10 (337) | 5.94 (4.23, 7.64) | 6.65 (5.09, 8.21) | 0.81 (-1.44, 3.06) | 0.481 |
| 2-back target, accuracy (%): 12-months | 10 (332) | 10 (337) | 7.82 (5.42, 10.22) | 7.39 (5.05, 9.74) | 0.05 (-2.79, 2.89) | 0.971 |
| d-prime: 6-months | 10 (328) | 10 (337) | -0.22 (-0.33, -0.12) | -0.11 (-0.21, -0.01) | 0.12 (-0.03, 0.26) | 0.109 |
| d-prime: 12-months | 10 (332) | 10 (337) | 0.61 (0.46, 0.77) | 0.58 (0.43, 0.74) | -0.01 (-0.20, 0.18) | 0.907 |

| Secondary outcomes | No of clusters (participants) | | Mean change from baseline (95% CI) | | Adjusted difference in follow-up ¹ | |
|---------------------------------------|-------------------------------|----------|------------------------------------|--------------------------|---|---------|
| | CON | INT | CON | INT | Coefficient (95% CI) | P value |
| 2-back non-target, RT (ms): 6-months | 10 (328) | 10 (337) | -61.33 (-83.33, -39.33) | -62.06 (-82.28, -41.84) | -1.96 (-31.07, 27.16) | 0.895 |
| 2-back non-target, RT (ms): 12-months | 10 (332) | 10 (337) | -83.28 (-116.48, -50.08) | -90.89 (-123.02, -58.76) | -18.58 (-55.68, 18.51) | 0.326 |
| 2-back target, RT (ms): 6-months | 10 (328) | 10 (337) | -53.55 (-73.24, -33.87) | -40.16 (-60.06, -20.26) | 11.10 (-16.31, 38.51) | 0.428 |
| 2-back target, RT (ms): 12-months | 10 (332) | 10 (337) | -57.83 (-88.40, -27.26) | -56.70 (-88.50, -24.91) | -7.10 (-42.13, 27.94) | 0.691 |

Note. RT = reaction time; ms = milliseconds; INT = intervention; CON = control; CI = confidence interval. Participants were included in the analysis if their overall mean flanker accuracy (i.e., congruent and incongruent) was higher than 50% and if their mean d-prime score was greater than 0 for the 1- and 2-back conditions. At baseline (37 cases), 6-month (13 cases) and 12-months (11 cases) flanker accuracy and reaction values were removed. At baseline (8 cases), 6-month (3 cases) and 12-months (2 cases) n-back accuracy and reaction values were removed.

¹Adjusted difference in secondary outcomes at 6- and 12-months = [(INT post-test mean minus INT baseline mean) minus (CON post-test mean minus CON baseline mean)].

P value adjusted for cluster effect, and randomisation pair.

Supplementary Table 10. Effect modification of flanker accuracy interference at 6-months

| Moderator | Interaction P value | Subgroup | Within group mean change (95% CI) (6-month minus baseline) | | Adjusted difference ¹ (INT v. CON) |
|----------------------|------------------------|--------------------------------------|---|------------------------|--|
| | | | CON | INT | Estimate |
| SES | 0.689 | High | -5.26 (-7.21 to -3.32) | -4.92 (-7.05 to -2.79) | 0.34 (-2.54 to 3.23) |
| | | Low or medium | -5.23 (-6.66 to -3.80) | -4.18 (-5.50 to -2.85) | 1.05 (-0.90 to 3.00) |
| Sex | 0.336 | Boy | -5.73 (-7.20 to -4.25) | -5.84 (-7.43 to -4.24) | -0.11 (-2.28 to 2.06) |
| | | Girl | -4.49 (-6.30 to -2.68) | -3.02 (-4.57 to -1.48) | 1.47 (-0.92 to 3.85) |
| Weight status | 0.392 | Healthy weight and underweight | -5.67 (-7.05 to -4.29) | -4.38 (-5.68 to -3.08) | 1.29 (-0.61 to 3.19) |
| | | Overweight and obese | -4.19 (-6.30 to -2.07) | -4.47 (-6.71 to -2.23) | -0.29 (-3.37 to 2.79) |
| Mental health status | 0.725 | Close to average and slightly raised | -5.09 (-6.39 to -3.79) | -4.43 (-5.66 to -3.21) | 0.66 (-1.13 to 2.44) |
| | | High to very high | -5.75 (-8.26 to -3.24) | -4.35 (-7.10 to -1.60) | 1.40 (-2.32 to 5.12) |
| CRF status | 0.494 | Health risk and needs improvement | -4.20 (-6.18 to -2.22) | -3.73 (-5.67 to -1.78) | 0.47 (-2.30 to 3.25) |
| | | Healthy fitness zone | -6.60 (-8.17 to -5.03) | -4.91 (-6.35 to -3.48) | 1.69 (-0.44 to 3.82) |

Note. INT = intervention; CON = control; SES = socio-economic status based on household postcode; CRF = cardiorespiratory fitness; CI = confidence interval

¹ Adjusted difference at 6-months = [(INT post-test mean minus INT baseline mean) minus (CON post-test mean minus CON baseline mean)].

Supplementary Table 11. Effect modification of flanker reaction time interference at 6-months

| Moderator | Interaction P value | Subgroup | Within group mean change (95% CI) (6-month minus baseline) | | Adjusted difference ¹ (INT v. CON) |
|----------------------|------------------------|--------------------------------------|---|-------------------------|--|
| | | | CON | INT | Estimate |
| SES | 0.151 | High | 0.66 (-5.12 to 6.43) | -6.99 (-13.35 to -0.62) | -7.64 (-16.24 to 0.95) |
| | | Low or medium | -2.51 (-6.76 to 1.75) | -2.56 (-6.50 to 1.38) | -0.06 (-5.86 to 5.75) |
| Sex | 0.243 | Boy | -1.46 (-5.86 to 2.95) | -6.99 (-11.76 to -2.22) | -5.53 (-12.02 to 0.96) |
| | | Girl | -0.98 (-6.40 to 4.44) | -0.77 (-5.40 to 3.85) | 0.21 (-6.92 to 7.33) |
| Weight status | 0.803 | Healthy weight and underweight | -0.30 (-4.42 to 3.82) | -3.26 (-7.15 to 0.62) | -2.97 (-8.63 to 2.69) |
| | | Overweight and obese | -3.75 (-10.04 to 2.54) | -5.35 (-12.02 to 1.32) | -1.60 (-10.77 to 7.57) |
| Mental health status | 0.880 | Close to average and slightly raised | -2.16 (-6.01 to 1.70) | -4.36 (-8.01 to -0.71) | -2.20 (-7.51 to 3.11) |
| | | High to very high | 1.62 (-5.83 to 9.07) | -1.53 (-9.69 to 6.64) | -3.14 (-14.20 to 7.91) |
| CRF status | 0.497 | Health risk and needs improvement | 0.99 (-4.88 to 6.87) | -2.69 (-8.49 to 3.10) | -3.69 (-11.94 to 4.57) |
| | | Healthy fitness zone | -3.76 (-8.44 to 0.92) | -3.85 (-8.12 to 0.43) | -0.08 (-6.42 to 6.26) |

Note. INT = intervention; CON = control; SES = socio-economic status based on household postcode; CRF = cardiorespiratory fitness; CI = confidence interval

¹ Adjusted difference at 6-months = [(INT post-test mean minus INT baseline mean) minus (CON post-test mean minus CON baseline mean)].

Supplementary Table 12. Effect modification of 2-back d-prime at 6-months

| Moderator | Interaction P value | Subgroup | Within group mean change (95% CI) (6-month minus baseline) | | Adjusted difference ¹ (INT v. CON) |
|----------------------|------------------------|--------------------------------------|---|------------------------|--|
| | | | CON | INT | Estimate |
| SES | 0.096 | High | -0.16 (-0.34 to 0.01) | -0.24 (-0.43 to -0.06) | -0.08 (-0.33 to 0.17) |
| | | Low or medium | -0.25 (-0.38 to -0.12) | -0.07 (-0.19 to 0.05) | 0.18 (0.01 to 0.35) |
| Sex | 0.547 | Boy | -0.23 (-0.36 to -0.10) | -0.07 (-0.21 to 0.07) | 0.16 (-0.03 to 0.35) |
| | | Girl | -0.20 (-0.37 to -0.04) | -0.13 (-0.27 to 0.01) | 0.07 (-0.14 to 0.29) |
| Weight status | 0.013 | Healthy weight and underweight | -0.15 (-0.27 to -0.03) | -0.15 (-0.26 to -0.03) | 0.00 (-0.16 to 0.17) |
| | | Overweight and obese | -0.38 (-0.56 to -0.19) | 0.03 (-0.17 to 0.23) | 0.41 (0.14 to 0.68) |
| Mental health status | 0.700 | Close to average and slightly raised | -0.22 (-0.34 to -0.11) | -0.10 (-0.21 to 0.01) | 0.12 (-0.03 to 0.28) |
| | | High to very high | -0.21 (-0.43 to 0.02) | -0.15 (-0.40 to 0.09) | 0.05 (-0.28 to 0.38) |
| CRF status | 0.108 | Health risk and needs improvement | -0.31 (-0.48 to -0.14) | 0.00 (-0.17 to 0.16) | 0.30 (0.07 to 0.54) |
| | | Healthy fitness zone | -0.18 (-0.31 to -0.04) | -0.12 (-0.24 to 0.00) | 0.06 (-0.12 to 0.24) |

Note. INT = intervention; CON = control; SES = socio-economic status based on household postcode; CRF = cardiorespiratory fitness; CI = confidence interval

¹ Adjusted difference at 6-months = [(INT post-test mean minus INT baseline mean) minus (CON post-test mean minus CON baseline mean)].

Supplementary Table 13: Intraclass correlation values expressing the similarity of individual observations within the same schools

| Outcome | School variance (Level 3) | Student variance (Level 2) | Error variance (Level 1) | ICC |
|----------------------------------|---------------------------|----------------------------|--------------------------|---------|
| Cardiorespiratory fitness (laps) | 15.8646 | 506.84 | 79.6235 | 0.02634 |
| Push-ups | 0.1522 | 61.1099 | 12.9233 | 0.00205 |
| Standing long jump | 42.1937 | 1187.69 | 116.03 | 0.03135 |
| Hair cortisol concentrations | 1.9148 | 0.1470 | 61.8951 | 0.02994 |
| Perceived stress | 0 | 22.2860 | 13.8743 | 0 |
| Internalising problems | 0 | 7.0399 | 2.9027 | 0 |
| Externalising problems | 0 | 7.8641 | 3.0561 | 0 |
| Steps per day (school hours) | 394434 | 542616 | 1332405 | 0.17380 |
| Steps per day (weekday) | 72613 | 2291422 | 1353060 | 0.01953 |
| Steps per day (weekend day) | 0 | 3100164 | 6255686 | 0 |
| MVPA mins/day (school hours) | 0.4643 | 44.0410 | 29.0537 | 0.00631 |
| MVPA mins/day (weekday) | 0 | 123.91 | 64.5551 | 0 |
| MVPA mins/day (weekend day) | 0 | 104.76 | 237.53 | 0 |

Note. MVPA = moderate-to-vigorous physical activity; ICC = intraclass correlation coefficient reflecting similarity of individual observations in the same classes. Some Level 3 variance estimates are 0 due to estimation of negligible or negative variances at Level 3, after accounting for variation at Levels 1 and 2, resulting in ICC estimates of 0.

Supplementary Table 14: Intraclass correlation values expressing the similarity of individual observations within the same classes

| Outcome | Class variance (Level 3) | Student variance (Level 2) | Error variance (Level 1) | ICC |
|----------------------------------|--------------------------|----------------------------|--------------------------|---------|
| Cardiorespiratory fitness (laps) | 31.0660 | 492.17 | 79.6281 | 0.05153 |
| Push-ups | 2.4018 | 58.9615 | 12.9248 | 0.03233 |
| Standing long jump | 65.3322 | 1164.61 | 116.03 | 0.04854 |
| Hair cortisol concentrations | 2.5205 | 0 | 61.5490 | 0.03934 |
| Perceived stress | 0 | 22.2858 | 13.8744 | 0 |
| Internalising problems | 0.07044 | 6.9716 | 2.9019 | 0.00708 |
| Externalising problems | 0 | 7.8651 | 3.0558 | 0 |
| Steps per day (school hours) | 477833 | 454396 | 1324904 | 0.21170 |
| Steps per day (weekday) | 148419 | 2212208 | 1356017 | 0.03993 |
| Steps per day (weekend day) | 0 | 3104281 | 6252301 | 0 |
| MVPA mins/day (school hours) | 3.7410 | 41.2045 | 28.9541 | 0.05062 |
| MVPA mins/day (weekday) | 3.8264 | 120.29 | 64.5087 | 0.02029 |
| MVPA mins/day (weekend day) | 0 | 104.76 | 237.52 | 0 |

Note. MVPA = moderate-to-vigorous physical activity; ICC = intraclass correlation coefficient reflecting similarity of individual observations in the same classes. Some Level 3 variance estimates are 0 due to estimation of negligible or negative variances at Level 3, after accounting for variation at Levels 1 and 2, resulting in ICC estimates of 0.