Indoor Air 2012; 22: 378–387
wileyonlinelibrary.com/journal/ina
Printed in Singapore. All rights reserved

INDOOR AIR
doi:10.1111/j.1600-0668.2012.00774.x

Effect of classroom air quality on students' concentration: results of a cluster-randomized cross-over experimental study

Abstract To assess the effect of indoor air quality as indicated by the median carbon dioxide (CO_2) level in the classroom on the concentration performance (CP) of students, a cross-over cluster-randomized experimental study was conducted in 20 classrooms with mechanical ventilation systems. Test conditions 'worse' (median CO_2 level on average 2115 ppm) and 'better' (median CO_2 level on average 1045 ppm) were established by the regulation of the mechanical ventilation system on two days in one week each in every classroom. Concentration performance was quantified in students of grade three and four by the use of the d2-test and its primary parameter 'CP' and secondary parameters 'total number of characters processed' (TN) and 'total number of errors' (TE). 2366 d2-tests from 417 students could be used in analysis. In hierarchical linear regression accounting for repeated measurements, no significant effect of the experimental condition on CP or TN could be observed. However, TE was increased significantly by 1.65 (95% confidence interval 0.42–2.87) in 'worse' compared to 'better' condition. Thus, low air quality in classrooms as indicated by increased CO_2 levels does not reduce overall short-term CP in students, but appears to increase the error rate.

D. Twardella[1], W. Matzen[2], T. Lahrz[3], R. Burghardt[3], H. Spegel[1], L. Hendrowarsito[1], A. C. Frenzel[4], H. Fromme[2]

[1]Department of Occupational and Environmental Health, Bavarian Health and Food Safety Authority, Munich, [2]Department of Toxicology and Chemical Safety, Bavarian Health and Food Safety Authority, Munich, [3]Berlin-Brandenburg State Laboratory, Department of Environmental Health Protection, Berlin, [4]Department of Psychology, University of Augsburg, Augsburg, Germany

Key words: Attention; Carbon Dioxide; Indoor Air; Students; School.

D. Twardella
Department of Occupational and Environmental Health
Bavarian Health and Food Safety Authority
Pfarrstrasse 3, 80538 Munich
Germany
Tel.: ++49-9131-6808 4249
Fax: ++49-9131-6808 4297
e-mail: dorothee.twardella@lgl.bayern.de

Received for review 26 October 2011. Accepted for publication 15 February 2012.

Practical Implications
This study could not confirm that low air quality in classrooms as indicated by increased CO_2 levels reduces short-term concentration performance (CP) in students; however, it appears to affect processing accuracy negatively. To ensure a high level of accuracy, good air quality characterized, for example, by low CO_2 concentration should be maintained in classrooms.

Introduction

In recent years, indoor environments in schools have come into the focus of discussion. In particular, the impact of indoor air quality on the attention and CP, achievements, well-being, and health of students has been discussed (Daisey et al., 2003; Haverinen-Shaughnessy et al., 2011; Mendell and Heath, 2005; Shendell et al., 2004).

Carbon dioxide (CO_2) has been commonly used as an indicator of indoor air quality. According to The German Working Group on Indoor Guidelines of the Federal Environment Agency and the States' Health Authorities, air quality can be regarded as 'harmless' if CO_2 levels are below 1000 ppm, 'elevated' if between 1000 and 2000 ppm, and 'hygienically unacceptable' if above 2000 ppm (Lahrz et al., 2008). This is in line with guidelines from other European countries (BMLFUW, 2006; UK Department of Education, 2006; NO-Folkehelseinstituttet 1996).

However, particularly in wintertime, increased CO_2 levels have been observed in classrooms. In a Bavarian measurement campaign in 91 classrooms, median CO_2 levels ranged between 598 and 4172 ppm (Fromme et al., 2008). In 25% of the classrooms, the median CO_2 level exceeded 2000 ppm and in 10%, 2700 ppm. Most classrooms rely on natural ventilation, and the cold temperature of outdoor air inhibits frequent window opening and causes accumulation of CO_2 in indoor air. High CO_2 levels in schools have been reported also from other countries (Daisey et al., 2003).

The relevance of air quality as indicated by CO_2 level in the classroom for the attention and concentration of the students could not be consistently shown yet. In a literature review from 2005, one single publication was identified, which analyzed the performance depending on indoor air quality (Mendell and Heath, 2005). A negative effect of low air quality on reaction time and performance was reported. More recent publications provide further indications for an effect of air quality on attention and concentration (Coley et al., 2007; Ribic, 2008; Wargocki and Wyon, 2007a,b). However, results are not sufficiently conclusive yet, because multiple testing as well as lack in blinding may have biased observed associations. On this background, we initiated the hereby reported RaBe (Raumluftqualität und das Befinden von Kindern – Indoor air quality and student experiences) study. The objective of the RaBe study was to assess the effect of indoor air quality as indicated by the CO_2 level in the classroom on the CP of students.

Methods

We conducted an experimental study with a cluster-randomized cross-over design. Data collection took place between November 2009 and April 2010. The RaBe study was approved by the Ethics Board of the Bavarian Chamber of Physicians, Germany and by the Bavarian Ministry of Education and Culture.

Participants

Classes of grade three and four (children usually aged 9–10 years) with a mechanical ventilation system in the classroom were eligible for participation. Primary schools in the German states, Bavaria and Berlin/Brandenburg, were approached by mail as well as personally, and consent for participation was obtained from the schools headmasters. Within each class, parents were informed about the study aims and procedures by written material as well as by an information evening in each school, and consent for data collection was obtained from parents of each student. All students of the identified classes were eligible for participation.

We included six schools in our study, five of which were located in the State of Bavaria and one close to Berlin. Most of the schools had been recently renovated or built. Per school two to six classes participated in the study. The classes consisted of 16–29 students of which between 67% and 100% agreed to participate. Overall, 417 students from 20 classes took part in the study (overall response rate for students 84%).

Experimental conditions

In each classroom, three experimental conditions were implemented:

- 'Usual': The mechanical ventilation was adjusted as usual. Window opening was allowed according to schools general regulations.
- 'Worse': The mechanical ventilation was down-regulated. Window opening was not allowed. It was aimed to reach a median CO_2 level of 2000–2500 ppm. Thus, conditions as can be found in wintertime in poorly ventilated classrooms were simulated.
- 'Better': The mechanical ventilation was up-regulated. It was aimed to reach a low median CO_2 level of < 1000 ppm. Window opening was not allowed. Thus, conditions as recommended by expert panels were simulated.

The reference period for the experimental condition was the beginning of the first class hour until the end of the d2-test in the fourth class hour (not counting breaks) and thus resulted in a exposure duration of typically 155 min. Experimental conditions were implemented on 2 days per week. Thus, the data collection period included 6 days in three consecutive weeks. For each class, those weekdays were preferably selected as test days, in which according to the timetable, all students usually stayed in the classroom all morning (at least from about 8 o'clock until 11 o'clock; i.e., German class hours 1–4). Classes from the same school were evaluated parallel during the same 3 weeks. Different schools were assessed consecutively.

Randomization

We randomized the unit 'school class,' because air quality could only be regulated classroom wise. In all classrooms, the experimental condition in the first week (the first two experimental days) was 'usual'. The sequence in the following 2 weeks – either worse/better or better/worse – was randomized. If the air quality in the classrooms within schools could be regulated separately, classrooms were arranged in pairs and randomization was performed within these pairs. In that way, in each pair, one classroom was assigned the sequence worse/better and the other the sequence better/worse. Otherwise, the sequence of conditions was determined for all classrooms within one school combined. Allocation was performed by random drawing of marked pieces of paper.

Outcomes

Concentration performance was assessed by the d2-test (Bates and Lemay, 2004; Brickenkamp, 2002). The d2-test is a one-page, paper-and-pencil test with 14 rows of the characters ‘d’ and ‘p’. The task is to mark as many target characters as possible (a ‘d’ with a total of two dashes placed above and/or below) per row in a limited time of 20 s. The test person is told verbally every 20 s to move on to the next line, leaving the previous line not fully examined. Based on the number of processed characters and the number of errors, specific outcome parameters can be determined. Those are ‘total number of characters processed’ (TN) as an indicator for processing speed and the ‘total errors’ (TE, defined as sum of incorrectly marked distractor characters plus left-out targets) as an indicator of accuracy, as well as ‘CP’ (defined as the number of correctly marked target characters minus incorrectly marked distractor characters) as an indicator for overall concentration. In RaBe, CP was used as a primary outcome. The d2-test was applied at each of the six experimental days in the fourth class hour. Trained study personnel instructed the students on how to fill in the form according to standard instruction. The students, teachers, and instructors were blinded with respect to the experimental condition on the day of the test. Non-participating children stayed in the classroom, and where given different tasks during the time, participating children filled out the questionnaire.

Other data collection

Sociodemographic data on the participating children were collected via a standardized questionnaire for parents. Characteristics of the schools and the classrooms were collected by a standardized documentation sheet, which was filled out by study assistants.

Measurement of air quality

During the 3 weeks of data collection, we documented the air quality in the respective classrooms. For this purpose, we placed an air-quality sensor (Klimawächter MF420-IR-CTF; J. Dittrich Elektronic GmbH & Co KG, Baden-Baden, Germany) in the middle of the classroom above the heads of the children. For each minute, the average room temperature (measurement range, 0–50°C), relative humidity (measurement range, 15–95%), and CO_2 level (measurement range, 0–3000 ppm) were recorded. The median CO_2 level, temperature, and relative humidity in a classroom on an experimental day were determined from the minute-by-minute measurements during the first four school hours (not counting the breaks) until the end of the d2-test. In case of gaps in data of maximal length of ten minutes (seven occasions), missing data were substituted by the average of the adjacent values. In case of longer gaps (nine occasions), missing values were substituted by measurement results from the second day with the same experimental condition in the respective classroom.

Statistical methods

We compared the distribution of baseline characteristics of classes and students in the two experimental arms (sequence usual–worse–better vs. sequence usual–better–worse). Grade and sex of the students were taken from school lists, other sociodemographic data from the parental questionnaire. We determined relative poverty based on the equivalent household net income, which was calculated from the household net income and the number and age of the household members weighted according to OECD (Bundesregierung 2005). If the equivalent household net income was $<60\%$ of the median equivalent net income in Germany, relative poverty was categorized as ‘yes’. Next, we described the CO_2 level in the classrooms depending on the experimental condition as well as the distribution of the parameters of the d2-test.

To determine the effect of the experimental condition on the test results (CP, TN, and TE), we employed three-level hierarchical linear models (Raudenbush and Bryk, 2002). Hierarchical linear models are an extension of the linear model, which is used in RaBe to account for the repeated measurements of students as well as for the correlation between students belonging to the same class. By defining test parameters on level one, students on level two, and classes on level three, the measured test parameters were examined within students within classes. The model assumes normally distributed errors on all three levels. Results of analysis are given on an additive scale.

First, we set up to a growth model to model the change in parameters with the repeated application of the d2-test (i.e., the learning effect). We allowed a linear and a quadratic term of growth. The inclusion of these terms as fixed effects on level one as well as random effects on level two and three was tested for significance at $P < 0.05$ based on the likelihood ratio test. In the resulting growth model, parameters to assess the impact of air quality were included as follows:

We followed three different approaches to test our hypothesis of air-quality effects on student performance in the d2. First, we used experimental condition (usual/worse/better) coded as dummy variables as predictors of the CP and included the complete data set. This analysis is in accordance with the design and comparable to an intention-to-treat approach in clinical studies. In an intention-to-treat approach, the complete data set has to be included in analysis and participants have to be analyzed as randomized irrespective of whether the treatment was in fact implemented or not.

As in some occasions the study protocol was not followed exactly, analysis was repeated using a reduced data set after exclusion of observations with deviations. Deviations from study protocol pertained to the completion of the d2-test (30 s instead of 20 s given for the completion of one of the 14 lines in the d2-test, student did not use the prescribed sign to mark relevant letters, student did not complete all 14 lines) as well as to test conditions (the whole class left the classroom for at least one hour during the morning, a part of the class left the classroom for at least one hour during morning, the class had physical education during the morning, the regulation of the ventilation system did not work). If the deviations from the study protocol were significantly associated with the test parameters in the hierarchical model and thus could introduce confounding, observations were excluded.

Finally, because of the variability of actual CO_2 levels within the study groups, in a third analysis, the association of the actual CO_2 median with students' concentration was estimated. In this analysis, instead of the study condition, the median CO_2 levels were introduced as a linear predictor in the model, as ventilation rate was found to be linearly related to student achievement (Haverinen-Shaughnessy et al., 2011). Analysis was based on the reduced data set as in the second analysis.

Below, we present results from three analyses: analysis 1 with complete data and experimental condition as predictor, analysis 2 with data cleaned for study protocol deviations and experimental condition as predictor, and analysis 3 with cleaned data and median CO_2 level as a predictor. In sensitivity analyses for the primary outcome CP, first analyses 1–3 were repeated after exclusion of observations from the 'usual' condition. Secondly, the effect of air quality on CP was determined in analyses 1–3 including only those observations in condition 'better', in which the CO_2 level was < 1000 ppm and those observations in condition 'worse' in which CO_2 level was > 2000 ppm. All hierarchical models were implemented in the software HLM 6.08.

Results

About half of the participating students were of grade 3 and about half were girls (Table 1). In seven percent of the students, parents reported dyslexia.

Classrooms were between 59 and 71 m^2 of size. The mean number of students of about 24 and the mean room volume of 215 m^3 resulted in a mean air volume of 9.2 m^3 available for each student (range, 6.9–14.9 m^3). Median room temperature on a test day ranged between 20.0 and 26.3°C (median = 23.6°C, 10% percentile 22.1°C, 90% percentile 25.1°C) and was similar on days with 'usual', 'worse', or 'better' condition. Median relative humidity in a classroom on a test day ranged between 15.0% and 42.5% (median = 31.5%, 10% percentile = 22.7%, 90% percentile = 39.4%) and was higher on days in 'worse' (median = 35.0%) than in 'better' condition (median = 26.9%).

Table 1 Sociodemographic background of the children included in the RaBe study

	Sequence of experimental condition			
	Usual–better–worse	Usual–worse–better	*P*-value[b]	Total
Grade				
Grade 3	140 (55%)	64 (39%)	<0.01	204 (49%)
Grade 4	114 (45%)	99 (61%)		213 (51%)
Sex				
Girls	115 (45%)	86 (53%)	0.14	201 (48%)
Boys	139 (55%)	77 (47%)		216 (52%)
Place of birth (missing = 27)[a]				
Germany	224 (95%)	148 (96%)	0.58	372 (95%)
Other	12 (5%)	6 (4%)		18 (5%)
Dyslexia (missing = 29)				
Yes	13 (6%)	13 (9%)	0.25	26 (7%)
No	222 (94%)	140 (92%)		362 (93%)
Hyperactivity (missing = 28)				
Yes/do not know	27 (11%)	8 (5%)	0.03	35 (9%)
No	208 (89%)	146 (95%)		354 (91%)
Single parent (missing = 29)				
Yes	32 (14%)	23 (15%)	0.70	55 (14%)
No	203 (86%)	130 (85%)		333 (86%)
Parental education (missing = 39)				
Low	32 (14%)	21 (14%)	0.91	53 (14%)
Average/High	199 (86%)	126 (86%)		325 (86%)
Relative poverty				
Yes	65 (26%)	33 (20%)	0.21	98 (24%)
No/unknown	189 (74%)	130 (80%)		319 (76%)

[a]Missing refers to missing data because of incomplete fill-out of study questionnaires. Subjects with missing data cannot be categorized.

[b]Cochrane Mantel–Haenszel *P*-value of overall association to compare the distribution of student characteristics in the two groups.

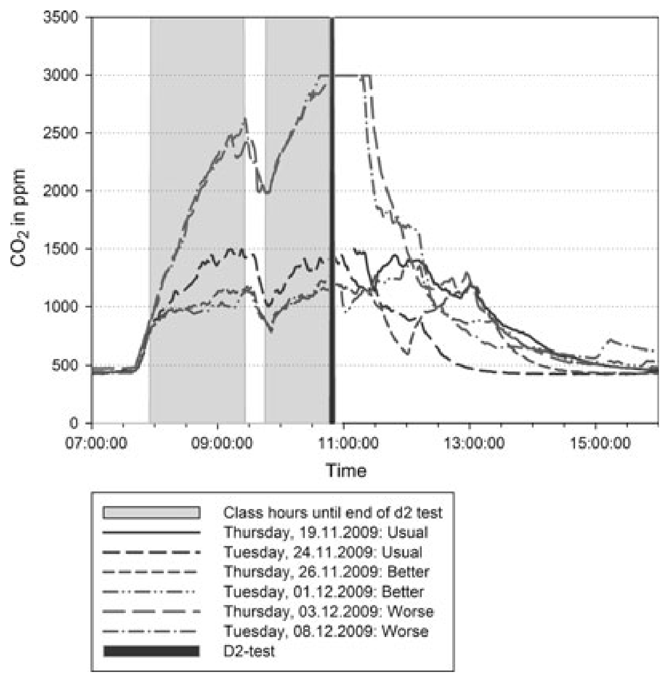

Fig. 1 Exemplary progression of CO_2 levels during the six experimental days in one of the classrooms of the RaBe study

Figure 1 shows an example of the progression of CO_2 level in one of the participating classrooms during the six experimental days. The general pattern of CO_2

Table 2 Distribution of the median CO_2 levels in the 20 classrooms of the RaBe study by experimental condition

	Usual	Worse	Better
N (classrooms)	40	40	40
Mean	1371	2115	1045
<1000 ppm	5	1	22
1000–<1500 ppm	23	1	17
1500–<2000 ppm	7	14	–
2000–2500 ppm	5	18	1
>2500 ppm	–	6	–

level in schools, which is typical for classrooms relying on natural ventilation, can be observed in our study with classrooms relying on mechanical ventilation, too. In the morning, CO_2 levels are low, increase during school hours, decrease during breaks, and decrease after students leave the school. In the experimental condition, ‘usual’ (blue lines) maximal CO_2 levels get close to 1500 ppm, in the ‘better’ condition (green lines) CO_2 levels were decreased, and in the ‘worse’ condition (red lines) increased. The gray shade marks the relevant time period from the beginning of the first school hour until the end of the d2-test. Looking at all classrooms, the fluctuation of CO_2 levels in the ‘better’ condition was smaller (average standard deviation of the mean CO_2 levels = 141) than in the ‘worse’ condition (average standard deviation = 559). The distribution of the minute-by-minute measurements in all classrooms is given in the Supporting information (Table S1).

Table 2 shows the distribution of the median CO_2 values during this relevant time period for all 20 participating classes for the three experimental conditions. It turned out to be difficult to regulate the mechanical ventilation system in a way that ensured the achievement of the planned CO_2 levels. Thus, only on 20 of 40 days in condition ‘worse’, the median CO_2 level was between 2000 and 2500 ppm, and only on 22 of 40 days in condition ‘better’, the CO_2 level was below 1000 ppm. However, on average, the median CO_2 level in the ‘worse’ condition (2115 ppm) was 1070 ppm higher than the CO_2 level in the ‘better’ condition (1045 ppm), which is significant ($P < 0.0001$) in the t-test.

In each data collection, round between 394 and 401 of the 417 students returned the d2-test, of which 391–397 could be used to calculate the test parameters, resulting in overall 2366 values for each d2-parameter. At the first occasion (round A), the mean values were 101 for CP, 270 for TN, and 13.7 for TE. During the following occasions, the mean values of CP and TN, but not TE, increased (Figure 2). Overall, values ranged from −75 to 298 for CP, 102–657 for TN, and 0–324 for TE.

In the hierarchical linear model, we could not observe a significant effect of the experimental condition on the primary parameter CP (Table 3). The CP

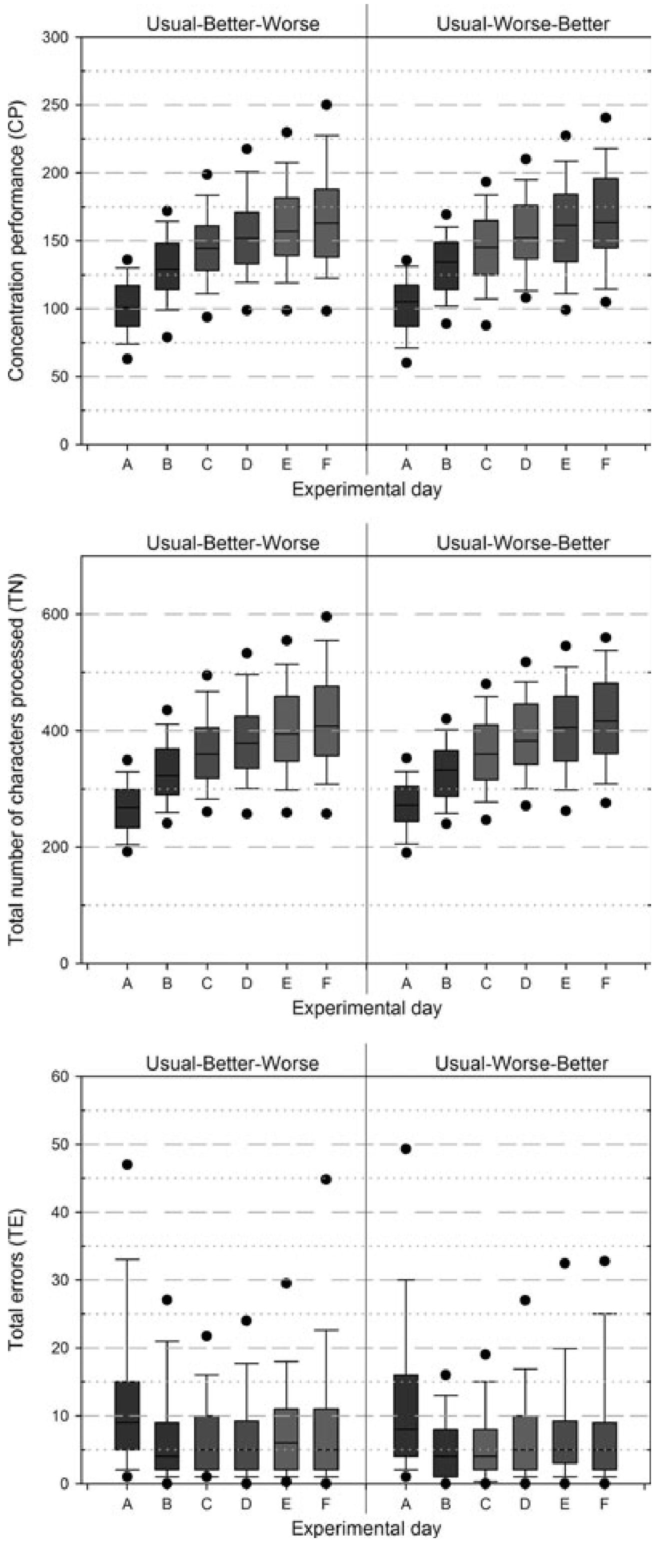

Fig. 2 Distribution of the test values during the six experimental days (round A to F) by sequence of experimental conditions. The boxplots show the median, the 25th and 75th percentile (boundary of the box), the 10th and 90th percentile (whiskers) and the 5th and 95th percentile (dots)

was decreased by 1.11 points at ‘worse’ in comparison with ‘better’ air quality in analysis 1; however, this difference was not statistically significant (95%

Table 3 Estimators of the effect of the experimental condition (analyses 1 and 2) and the actual CO_2 level (analysis 3) on the d2-test parameters with 95% confidence intervals

	Number of observations in analysis Reduction in test parameter (95% CI)		
	Analysis 1	Analysis 2	Analysis 3
Concentration performance			
Worse compared to better air quality	*N* = 2366 −1.11 (−2.44; 0.22)	*N* = 1962 −0.55 (−1.83; 0.73)	
Median CO_2 (per 1000 ppm)			*N* = 1962 −0.76 (−1.86; 0.34)
Total number of characters processed (TN)			
Worse compared to better air quality	*N* = 2366 −0.88 (−3.84; 2.08)	*N* = 2038 −0.11 (−3.23; 3.01)	
Median CO_2 (per 1000 ppm)			*N* = 2038 −0.88 (−3.46; 1.70)
Total errors (TE)			
Worse compared to better air quality	*N* = 2366 1.34 (−0.03; 2.70)	*N* = 2254 1.65 (0.42; 2.87)	
Median CO_2 (per 1000 ppm)			*N* = 2254 1.19 (0.30; 2.07)

confidence interval CI −2.44 to 0.22). No significant effect could be found, if observations were excluded (analysis 2) or if the actual median CO_2 level instead of the experimental condition was tested (analysis 3). In sensitivity analysis after exclusion of observations from round A and B, effect estimators were even lower and also not statistical significant (analysis 1: 0.45, 95% CI −1.96 to 1.07; analysis 2: 0.35, 95% CI −1.76 to 1.06; analysis 3: 0.37, 95% CI −1.56 to 0.82). If only observations in condition 'better' were included, in which the CO_2 level was <1000 ppm and only those observations in condition 'worse' in which CO_2 level was >2000 ppm, again, no significant effect of experimental condition on CP was found (data not shown).

Similarly, no significant effect of experimental condition or median CO_2 level on TN could be observed (Table 3). With respect to TE, though, a significant result was produced. In analysis 1, using all observations, the number of errors was increased by 1.34 (95% CI −0.03 to 2.70) in the worse air-quality condition compared to better air quality. If observations with deviation from the protocol were excluded, the effect became stronger and statistical significance was reached. In analysis 2, the number of errors was significantly increased by 1.65 (95% CI 0.42–2.87) in case of worse air quality. Also, the TE was increased with increased median CO_2 level. With an increase in median CO_2 by 1000 ppm, TE increased by 1.19 (95% CI 0.30–2.07).

Discussion

We could recruit 20 school classes in the RaBe study, of which 417 students participated in data collection. The regulation of the mechanical ventilation system to reach the planned CO_2 levels in the classrooms turned out to be challenging. However, a significant gradient of 1070 ppm CO_2 between 'worse' and 'better' experimental conditions could be achieved, indicating a potentially relevant difference in air quality. The hypothesis that worse air quality as indicated by increased CO_2 level causes a reduction in CP in students could not be confirmed with our study. While the estimators for the effect air quality on CP did not reach statistical significance, all point in the same direction, that is, low performance at worse air quality. Furthermore, results indicate a negative effect of worse air quality on accuracy.

In the past, few studies have been conducted on the CP of students and their relation to air quality as indicated by CO_2 levels. In a review from the year 2005, only one study was included which evaluated students' performance depending on air quality (Mendell and Heath, 2005). In this experimental study, a negative effect of high CO_2 level on reaction time and performance could be observed. Since then, seven other studies on this issue have been published.

In an observational study from the USA, the air quality with closed windows and active ventilation system was measured in classroom of classes grade 5 during one school day and the ventilation rate deduced. Standardized test scores based on math and reading skills were obtained for the students. In the pilot analysis of 50 school classes, the association between the ventilation rate and test results was significant only at $P < 0.1$ but not at $P < 0.05$ (Shaughnessy et al., 2006). In the main study, using only data of those classes, in which ventilation rates were below recommended guidelines ($N = 87$), a significant association between the ventilation rate and the results of the math and reading tests could be observed (Haverinen-Shaughnessy et al., 2011). This study provides indication of the relevance of indoor air quality for long-term student achievement as indicated by standardized tests. However, because of the observational design, the validity is limited.

The five remaining studies are experiments, in which students were tested at different ventilation rates resulting in different CO_2 levels. Naturally, only short-term performance at the time of the specific air quality can be measured in such a design. In Denmark, two experimental studies have been undertaken in two classrooms of one school with students of age ten to twelve (Wargocki and Wyon, 2007a,b). Both were cross-over trials in which CO_2 levels were regulated by the ventilation system. However, additional window opening was allowed. The performance of the students was assessed by seven exercises exemplifying different aspects of schoolwork. In the first experimental study, the difference in CO_2 levels between study conditions was low (on average 1270 with low and 920 ppm with high ventilation), but still a significant higher speed of work with high ventilation was found for five of the seven exercises (Wargocki and Wyon, 2007b). No effect on accuracy was found. In the second study, because of missing data, only four of the seven exercises could be analyzed (Wargocki and Wyon, 2007a). For none of those exercises, a significant influence of ventilation was observed. The difference in CO_2 level in this trial was even lower (about 775 vs. 1000 ppm).

In a study from England, computer-based tests were conducted on 10 days in one class with students aged ten to eleven years (Coley et al., 2007). Poor air quality with mean CO_2 levels of 2909 ppm was reached by restriction of window opening on 5 days, while on the other 5 days by opening of the windows mean CO_2 levels of 690 ppm were achieved. For three of the 13 parameters of cognitive function, a significant influence of air quality was observed. At good air quality, reaction time was reduced and attention increased, while at poor air-quality calmness was increased. It remains unclear from the publication, whether the learning effect was considered in the analysis. Furthermore, because study conditions were established by window opening, students and teachers were not blinded with respect to the study condition, which could have an impact on test results.

In a second publication from the UK, preliminary results of an experimental cross-over trial in one school were reported (Bako-Biro et al., 2007). A direct air supply system through the window was used to establish two study conditions: (i) provide outdoor air (mean CO_2 593–783 ppm); (ii) recirculate the classroom air (mean CO_2 1638–4093 ppm). Students' performance was tested with a 40-min paper-based test (reading comprehension, addition, subtraction). No effect of study condition on reading comprehension or subtraction, but a significant improvement of addition with provision of fresh air was observed.

Lastly, an experimental cross-over study in six classes in Switzerland was conducted, and performance assessed with the d2-test in students aged 15–16 years (Ribic, 2008). Air quality was influenced by regulations on window opening and CO_2 levels of 600–800 ppm for good air quality and of at least 3000 ppm for low air quality established. A significantly reduced CP was observed at low air quality. However, students and teachers were not blinded with respect to the study condition. Furthermore, it remains unclear whether the tests were completed under the same conditions, for example, at the same time of day.

The above-mentioned studies mostly reported a significant effect of air quality on at least some parameters of students' performance. However, in some studies, methodological limitations confine the interpretation. The lack of blinding may have caused false-positive associations. Parallel testing of multiple performance parameters without correction of the *P*-value weakens the interpretation of statistical significance. Furthermore, the correlations between students coming from the same class have not been accounted for in the statistical analysis of most of the mentioned studies. To neglect correlation between observations – a clustered design – leads to an underestimation of variance and thus can cause false statistical significance (Donner and Klar, 2000). In our study, we tried to avoid these methodological problems. We included a relatively large number of classes and students and were able to account for correlations between classes in statistical analysis. Students, teachers, and test instructors were blinded with respect to the experimental condition. We defined CP as our primary outcome to avoid multiple testing. Results of secondary outcome analysis are interpreted as explorative results.

With our RaBe study, we were not able to confirm the results of former studies, which found a decreased CP of children exposed to low air quality as indicated by high CO_2 levels. The following limitations have to be accounted for when interpreting our study results. Mechanical ventilation in classrooms is very seldom in Germany, and only recently, if schools are newly built or if major renovations are undertaken, ventilation systems are installed. Thus, although we aimed for 24 classes with 600 students, only 20 classes with 417 students took part. With a larger sample size, the effect estimator might have become statistically significant. However, even if statistically significant, it might be questionable, whether a diminishment by 1% (effect estimator 1.11 points, mean value in round A 101 points) would be regarded as a relevant change.

It was not always possible to adapt the timetable and routine in schools to the optimal study design. Thus, we could not avoid that in some cases classes had physical education or left the classroom in the morning before the test. We tried to compensate for these deviations from study protocol by running analysis 2. However, results did not change substantially. Still, school is not a laboratory, and it is difficult to standardize all possible influential factors.

Confounding is an unlikely explanation for the observed results. First of all, one has to consider that RaBe is a cross-over study, and thus, each student serves as its own control. Thus, in the statistical model, confounding of the effects size by any characteristic of the student is not possible. However, a statistical interaction cannot be excluded, which would be interpreted as a differential effect of air quality depending on student characteristics. As the model that was used for RaBe was based on the assumption that the effect of air quality on CP does not depend on the characteristic of the students, it does not account for interactions.

Secondly, RaBe was a randomized study. Randomization per se is a method to control confounding by design (Rothman et al., 2008). If randomization is successful, confounding by know as well as unknown factors is prevented and thus adjustment in statistical analysis not needed. However, the number of randomized units (classes) is relatively low and thus some residual confounding possible.

To tackle the issue of residual confounding, the following steps were taken: First of all, because of the obvious strong enhancement of test parameters with repeated testing, we assessed the effect of experimental condition within a growth model, which captured this change over time. Thus, all reported results are adjusted by the 'learning effect'. Secondly, for the primary outcome CP, we conducted adjusted analysis. Sociodemographic characteristics as well as room temperature and relative humidity were assessed for relevance in the growth model, and the effect of experimental condition on CP was tested in an model adjusted for those factors showing $P < 0.1$ in the growth model. In this model, only the factor 'single parent' showed a significant interaction. While in the adjusted model there was a significant reduction of CP in 'lower air quality' (−1.85, 95% CI −3.38 to −0.33), in children with a single parent this was changed to an increase by 2.83. There is no obvious explanation for this interaction and chance result because multiple testing is possible.

The CO_2 measurement range had an upper detection limit of 3000 ppm, which was reached on eight of the 120 test occasions. The exceedance lasted for 1, 2, 3, 6, 7, 18, 39, or 49 min, respectively, on these eight occasions. In none of the classes, the percentage of the minute-by-minute CO_2 values on one test day, which were above the detection limit reached 50%. Thus, the calculation of median CO_2 levels was not affected by this measurement limitation.

We were not able to achieve the exact CO_2 levels as planned by regulation of the mechanical ventilation system. However, first, we could show that on average CO_2 levels were considerably higher in the 'worse' condition than in the 'better' condition. Secondly, in sensitivity analysis, we run a model in which only those observations were included, in which the specifications were achieved. Still in this analysis, no effect of experimental condition on CP could be observed. Thus, it seems implausible that the difficulties in achieving specific CO_2 levels were responsible for the absence of the effect.

Although students, teachers, and test instructors were blinded with respect to the experimental condition, we cannot exclude that at least some subjects could have become aware of the experimental condition during the test period because of body odors and general stuffiness in the room. Thus, the reported association might at least partly be due to insufficient blinding.

The RaBe study differs with respect to the range of observed CO_2 levels from other studies. Effects of CO_2 levels might only become evident when comparing more distinct groups or air quality or achieving CO_2 levels far below 1000 ppm.

In RaBe, student performance was characterized with the d2-test. In contrast to usual school examinations, the test does not require specific skills but aims to assess the aspect of concentration as a basic prerequisite for the provision of any achievement (Brickenkamp, 2002). In several studies, a positive correlation between the CP derived from the d2-test and intelligence and achievement motivation has been observed (see for example Romainczyk, 2008 or Schaal, 2004). It has also been shown that parameters of the d2-test correlate with school marks (Brickenkamp, 2010). In some of the other published studies, school work was used to characterize student's performance. While performance at school work is a more direct measurement of a relevant outcome, which can be obtained in a usual school situation, the knowledge and skills of the students will impact the performance to a larger extent particularly if the same examination will be repeated at different times. Both studies using tightly controlled tests as in our study (Coley et al., 2007; Ribic, 2008) as well as studies using school examinations to characterize student performance (Wargocki and Wyon, 2007a) obtained a significance influence of air quality. There are no indications for a differential effect. The d2-test does assess concentration only. Other possible effects of air quality such as an increase in health symptoms and resulting absenteeism are not covered.

In our study, air quality is described by carbon dioxide concentration and not by ventilation rates. By design, carbon dioxide levels were used to define experimental groups. Carbon dioxide is a well-accepted indicator of air quality, and parameters of ventilation engineering are of less importance in a country such as Germany with the majority of schools relying on natural ventilation. Furthermore, as in RaBe classrooms serve as their own control and CO_2 generation can be assumed constant within classrooms, differences in ventilation rates in a classroom at different experimental conditions will highly correlate with differences in CO_2 levels. Even if ventilation rates were derived, results of analyses 1 and 2 would not change because the predictor variable was experimental group.

Using the median CO_2 levels, the situation in the classrooms in RaBe can be easily compared to real-life situation that is observed in classrooms in Germany as has been described in the past (Fromme et al., 2008). CO_2 is a well-known indicator of air quality and has been related to health effects in the past (see for example Shendell et al., 2004; Seppänen et al., 1999).

The reason for the absence of a significant effect on the CP in our study does not become clear. We can speculate, but not prove with data, that motivation of students might have played a role. In RaBe, we observed that the children looked forward to the test and were highly motivated. The test was an interruption from the usual school work, and the children may have perceived it as a welcome diversion. In addition, the d2-test lasts only about five minutes and thus requires only short-term attention. It could be that for a short period of time children can activate resources even if exposed to low air quality. This would explain why we could not observe an effect on the CP. It cannot be excluded that in longer tests low air quality eventually diminishes attention and CP.

Furthermore, while we could not observe an impact on the primary outcome CP, our data do suggest a decrease in the secondary outcome accuracy if children are exposed to poor air quality. In comparison with the mean total error of between 7.8 and 13.7, an increase in 1 to 1.5 points is an increase by about 10% and thus of relevance. The association is significant if looking at the effect of the experimental condition as well as looking at the effect of the CO_2 level itself. In line with the previous interpretation, one could argue that children are able to achieve a normal CP for a short time even if exposed to poor air quality, but while they are able to sustain the processing speed they risk more errors. Thus, tasks that require high precision might be more strongly affected by poor air quality. However, this result is in contradiction with the study of Wargocki et al. in which increased ventilation increased the speed of work but did not have any effect on accuracy (Wargocki and Wyon, 2007b). The reason for this contradictory result remains unclear.

Conclusion

We were not able to show a negative effect of air quality, which was similar to that observed at wintertime in classrooms relying on natural ventilation, on the CP of students. Secondary analysis suggests an effect of air quality on the processing accuracy. Our results are in conflict with earlier studies, which suggested the presence of an effect on concentration. The causes of the conflicting results remain unclear. It cannot be excluded that significant results of earlier studies are at least partly due to methodological limitations. While we tried to avoid these limitations in RaBe, bias as a result of varying circumstances in schools that are difficult to standardize cannot be ruled out. One possible explanation could be that poor air quality does not impair short-term CP (as measured by the d2-test in our study) but only shows its effect in longer concentration efforts. Furthermore, air quality might be more relevant for the precision than for the processing speed. Further research is needed to clarify the relevance of indoor air quality on long-term CP of students.

Acknowledgements

We wish to thank the participating schools, the teachers, and students for their support of the study as well as Ramona Grahle, Annemarie Hiergeist, Vera Hoffmann, Petra Kaplan, Petra Panenka, Annette Mangstl, Evelyn Schmidt, Ute Warmuth, and Simone Zyzik-Zinn for their engagement in the data collection and data management and Angelika Schwaiger for assistance in organizational matters.

Competing interests

None.

Funding

The study was funded by the Deutsche Bundesstiftung Umwelt (DBU), Az. 27549 – 25. The DBU was not involved in the study design; in the collection, analysis, and interpretation of the data; in the writing of the report; or in the decision to submit the paper for publication. The manuscript reflects the perception and opinion of the contractor. It does not represent the opinion of the DBU.

Supporting Information

Additional Supporting Information may be found in the online version of the article:

Table S1 Distribution of the minute-by-minute measurements of carbon dioxide concentration in each classroom during the six 4-hour test periods.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

References

Bako-Biro, Z., Kochhar, N., Clements-Croome, D.J., Awbi, H.B. and Williams, M. (2007) Ventilation rates in schools and learning performance. *Proceedings of Clima 2007 WellBeing Indoors*, June 2007, Helsinki.

Bates, M.E. and Lemay, E.P. Jr (2004) The d2 Test of attention: construct validity and extensions in scoring techniques, *J. Int. Neuropsychol. Soc.*, **10**, 392–400.

BMLFUW – Bundesministerium für Land und Forstwirtschaft, Umwelt und Wasserwirtschaft (2006) Physikalische Faktoren Kohlenstoffdioxid als Lüftungsparameter, Wien.

Brickenkamp, R. (2002) *Test d2. Aufmerksamkeits- und Belastungstest. Manual*, 9th edn. Göttingen, Hogrefe Verlag.

Brickenkamp, R., Schmidt-Atzert, L. and Liepmann, D. (2010) *Test d2 – Revision: Aufmerksamkeits- und Konzentrationstest*. Göttingen, Hogrefe Verlag.

Bundesregierung (2005) *Lebenslagen in Deutschland – Zweiter Armuts- und*

Reichtumsbericht. Köln, Bundesanzeiger Verlagsgesellschaft.

Coley, D.A., Greeves, R. and Saxby, B.K. (2007) The effect of low ventilation rates on the cognitive function of a primary school class, *Int. J. Vent.*, **6**, 107–112.

Daisey, J.M., Angell, W.J. and Apte, M.G. (2003) Indoor air quality, ventilation and health symptoms in schools: an analysis of existing information, *Indoor Air*, **13**, 53–64.

Donner, A. and Klar, N. (2000) *Design and Analysis of Cluster Randomization Trials in Health Research*, 1st edn. London, Arnold.

Fromme, H., Heitmann, D., Dietrich, S., Schierl, R., Korner, W., Kiranoglu, M., Zapf, A. and Twardella, D. (2008) Raumluftqualität in Schulen – Belastung von Klassenräumen mit Kohlendioxid (CO2), flüchtigen organischen Verbindungen (VOC). Aldehyden, Endotoxinen und Katzenallergenen [Air quality in schools - classroom levels of carbon dioxide (CO2), volatile organic compounds (VOC), aldehydes, endotoxins and cat allergen], *Gesundheitswesen* **70**, 88–97.

Haverinen-Shaughnessy, U., Moschandreas, D.J. and Shaughnessy, R.J. (2011) Association between substandard classroom ventilation rates and students' academic achievement, *Indoor Air*, **21**, 121–131.

Lahrz, T., Bischof, W. and Sagunski, H. (2008) Gesundheitliche Bewertung von Kohlendioxid in der Innenraumluft. Mitteilungen der Ad-hoc-Arbeitsgruppe Innenraumrichtwerte der Innenraumlufthygiene-Kommission des Umweltbundesamtes und der Obersten Landesgesundheitsbehörden, *Bundesgesundheitsbl Gesundheitsforsch Gesundheitsschutz*, **51**, 1358–1369.

Mendell, M.J. and Heath, G.A. (2005) Do indoor pollutants and thermal conditions in schools influence student performance? A critical review of the literature, *Indoor Air*, **15**, 27–52.

NO-Folkehelseinstituttet (1996). Anbefalte faglige normer for inneklima. http://www.fhi.no/dav/249C03CE-C6614E87862368DA175E7A31.pdf.

Raudenbush, S.W. and Bryk, A.S. (2002) *Hierarchical Linear Models. Applications and Data Analysis Methods*, 2nd edn, Thousand Oaks, London, New Delhi, Sage Publications.

Ribic, W. (2008) Nachweis des Zusammenhangs zwischen Leistungsfähigkeit und Luftqualität, *HLH* Lüftung/Klima – Heizung/Sanitär – Gebäudetechnik, **59**, 43–46.

Romainczyk, B. (2008) *Bearbeitungsstrategien in Konzentrationstests*, Diploma Thesis.

Rothman, K.J., Greenland, S. and Lash, T.L. (2008) *Modern Epidemiology*, 3rd edn, Philadelphia, Lippincott Williams & Wilkins.

Schaal, S. (2004) *Leistungsmotivation: eine Untersuchung zur Validität verschiedener Verfahren*, Diploma Thesis, Marburg.

Seppänen, O.A., Fisk, W.J. and Mendell, M.J. (1999) Association of ventilation rates and CO_2 concentrations with health and other responses in commercial and institutional buildings, *Indoor Air*, **9**, 226–252.

Shaughnessy, R.J., Haverinen-Shaughnessy, U., Nevalainen, A. and Moschandreas, D. (2006) A preliminary study on the association between ventilation rates in classrooms and student performance, *Indoor Air*, **16**, 465–468.

Shendell, D.G., Prill, R., Fisk, W.J., Apte, M.G., Blake, D. and Faulkner, D. (2004) Associations between classroom CO_2 concentrations and student attendance in Washington and Idaho, *Indoor Air*, **14**, 333–341.

UK Department of Education (2006) Ventilation of School Buildings. Building Bulletin 101. http://www.education.gov.uk/schools/adminandfinance/schoolscapital/buildingsanddesign/a0058229/ventilation-and-indoor-air-quality-in-schools-building-bulletin-101.

Wargocki, P. and Wyon, D.P. (2007a) The effects of moderately raised classroom temperatures and classroom ventilation rate on the performance of schoolwork by children (1257-RP), *ASHRAE HVAC&R Res.*, **13**, 193–220.

Wargocki, P. and Wyon, D.P. (2007b) The effects of outdoor air supply rate and supply air filter condition in classrooms on the performance of schoolwork by children (1257-RP), *ASHRAE HVAC&R Res.*, **13**, 165–191.