Fluid intelligence (Gf) is a complex human ability that allows us to adapt our thinking to a new cognitive problem or situation (1). Gf is critical for a wide variety of cognitive tasks (2), and it is considered one of the most important factors in learning. Moreover, Gf is closely related to professional and educational success (3–6), especially in complex and demanding environments. Although performance on tests of Gf can be improved through direct practice on the tests themselves, there is no evidence that training on any other regimen yields increased Gf in adults. Furthermore, there is a long history of research into cognitive training showing that, although performance on trained tasks can increase dramatically, transfer of this learning to other tasks remains poor. Here, we present evidence for transfer from training on a demanding working memory task to measures of Gf. This transfer results even though the trained task is entirely different from the intelligence test itself. Furthermore, we demonstrate that the extent of gain in intelligence critically depends on the amount of training: the more training, the more improvement in Gf. That is, the training effect is dosage-dependent. Thus, in contrast to many previous studies, we conclude that it is possible to improve Gf without practicing the testing tasks themselves, opening a wide range of applications.

Fluid intelligence (Gf) refers to the ability to reason and to solve new problems independently of previously acquired knowledge. Gf is critical for a wide variety of cognitive tasks, and it is considered one of the most important factors in learning. Moreover, Gf is closely related to professional and educational success, especially in complex and demanding environments. Although performance on tests of Gf can be improved through direct practice on the tests themselves, there is no evidence that training on any other regimen yields increased Gf in adults. Furthermore, there is a long history of research into cognitive training showing that, although performance on trained tasks can increase dramatically, transfer of this learning to other tasks remains poor. Here, we present evidence for transfer from training on a demanding working memory task to measures of Gf. This transfer results even though the trained task is entirely different from the intelligence test itself. Furthermore, we demonstrate that the extent of gain in intelligence critically depends on the amount of training: the more training, the more improvement in Gf. That is, the training effect is dosage-dependent. Thus, in contrast to many previous studies, we conclude that it is possible to improve Gf without practicing the testing tasks themselves, opening a wide range of applications.

Improving fluid intelligence with training on working memory

Susanne M. Jaeggi**, Martin Buschkuehl**, John Jonides*, and Walter J. Perrig†

*Department of Psychology, University of Michigan, East Hall, 530 Church Street, Ann Arbor, MI 48109-1043; and †Department of Psychology, University of Bern, Muesmattstrasse 45, 3012 Bern, Switzerland

Edited by Edward E. Smith, Columbia University, New York, NY, and approved March 18, 2008 (received for review February 7, 2008)

Fluid intelligence (Gf) is a complex human ability that allows us to adapt our thinking to a new cognitive problem or situation (1). Gf is critical for a wide variety of cognitive tasks (2), and it is considered one of the most important factors in learning. Moreover, Gf is closely related to professional and educational success (3–6), especially in complex and demanding environments. Although performance on tests of Gf can be improved through direct practice on the tests themselves, there is no evidence that training on any other regimen yields increased Gf in adults. Furthermore, there is a long history of research into cognitive training showing that, although performance on trained tasks can increase dramatically, transfer of this learning to other tasks remains poor. Here, we present evidence for transfer from training on a demanding working memory task to measures of Gf. This transfer results even though the trained task is entirely different from the intelligence test itself. Furthermore, we demonstrate that the extent of gain in intelligence critically depends on the amount of training: the more training, the more improvement in Gf. That is, the training effect is dosage-dependent. Thus, in contrast to many previous studies, we conclude that it is possible to improve Gf without practicing the testing tasks themselves, opening a wide range of applications.

Fluid intelligence (Gf) is a complex human ability that allows us to adapt our thinking to a new cognitive problem or situation (1). Gf is critical for a wide variety of cognitive tasks (2), and it is considered one of the most important factors in learning. Moreover, Gf is closely related to professional and educational success (3–6), especially in complex and demanding environments. Although performance on tests of Gf can be improved through direct practice on the tests themselves, there is no evidence that training on any other regimen yields increased Gf in adults. Furthermore, there is a long history of research into cognitive training showing that, although performance on trained tasks can increase dramatically, transfer of this learning to other tasks remains poor. Here, we present evidence for transfer from training on a demanding working memory task to measures of Gf. This transfer results even though the trained task is entirely different from the intelligence test itself. Furthermore, we demonstrate that the extent of gain in intelligence critically depends on the amount of training: the more training, the more improvement in Gf. That is, the training effect is dosage-dependent. Thus, in contrast to many previous studies, we conclude that it is possible to improve Gf without practicing the testing tasks themselves, opening a wide range of applications.
To investigate whether training on working memory leads to transfer to Gf, we conducted four individual experiments all using a newly developed training paradigm consisting of a very demanding working memory task, illustrated in Fig. 1. In this task, participants saw two series of stimuli that were synchronously presented at the rate of 3 s per stimulus. One string of stimuli consisted of single letters whereas the other consisted of individual spatial locations marked on a screen. The task was to decide for each string whether the current stimulus matched the one that was presented n items back in the series. The value of n varied from one block of trials to another, with adjustments made continuously for each participant based on performance. As performance improved, n incremented by one item; as it worsened, n decremented by one item. Thus, the task changed adaptively so that it always remained demanding, and this demand was tailored to individual participants. This form of training engaged processes required for the management of two simultaneous tasks; it engaged executive processes required for each task; and it discouraged the development of task-specific strategies and the engagement of automatic processes because of the variation in n and because of the inclusion of two different classes of stimuli.

The aim of the training intervention was the investigation of the effects of training on the working memory task and its impact on Gf. We accomplished this investigation by pretesting participants on a measure of Gf and then posttesting them on another form of this measure. Because we hypothesized that any alteration of the processing system would take some time to be effective, an important difference among the four experiments was the number of training sessions between pre- and posttests, ranging from 8 to 19 sessions. To control for mere retest effects, the performance of the trained groups was compared with control groups who were also assessed on Gf, but who were not trained between the two testing sessions.

Results

Analyses of the training functions revealed that all four training groups improved in their performance on the working memory task comparably (Fig. 2). What interests us most, however, is the dramatic improvement on the test of Gf in the trained groups (Fig. 3a). Although the gain in the control groups was also significant, presumably because of retest effects (t(34) = 2.08; P < 0.05; Cohen’s d = 0.25), the improvement in the groups that received the apparent benefit of training was substantially superior (t(33) = 5.53; P < 0.001; Cohen’s d = 0.65), which was confirmed by the significant group × test-session interaction (F(1,67) = 5.27; P < 0.05; η² = 0.07). A subsequent analysis of the gain scores (posttest minus pretest) as a function of training time (8, 12, 17, or 19 days) showed that transfer to fluid intelligence varied as a function of training time (F(3,30) = 9.25; P < 0.001; η² = 0.48; Fig. 3b). Analyses of covariance (ANCOVA) with the factor group (trained vs. control), the posttest scores as the dependent variable, and the pretest scores as the covariate revealed a trend for group differences after 12 days (F(1,19) = 1.93; P = 0.09; η² = 0.09), and statistically significant group differences after 17 (F(1,13) = 4.65; P < 0.05; η² = 0.26), and 19 training days (F(1,12) = 4.53; P < 0.05; η² = 0.27). Post hoc analyses (Gabriel’s procedure; two-tailed) for the training group revealed significant differences between the following groups: 8 vs. 17 days (P < 0.01); 8 vs. 19 days (P < 0.001); and 12 vs. 19 days (P < 0.1). There was a trend for a difference between 12 and 17 days (P = 0.06). These analyses indicate that the gain in fluid intelligence was responsive to the dosage of training.

It is important to note that the gain in Gf is strictly training-related and not due to preexisting individual differences in intelligence or working memory. There was an effect of training (F(1,68) = 6.38; P < 0.05; η² = 0.09) irrespective of initial Gf as shown by dividing the sample into high and low performers in Gf at pretest (by median split). However, there was also a main effect of performance group (F(1,68) = 4.56; P < 0.05; η² = 0.07), showing that participants with initially lower Gf generally showed even larger gains in Gf. There was no significant performance group by training-gain interaction (F < 1).

Further, the gain in Gf was not dependent on initial working memory capacity as assessed by either pretest performance in a digit-span task (F(1,68) < 0.3) or a reading-span task (F(1,52) < 0.1; note that reading span was not assessed in the 8-day group). Therefore, our cognitive training proved to be useful for all participants and had no adverse effects for participants with high initial working memory capacity.

Additional analyses showed that there was training-related transfer on the digit-span task (group × session interaction:
capacity as measured by a performance increase in both the
mained intact after controlling for the gain in working memory
over training time. \((Gf)\text{span task, in which both groups improved equally in the posttest}
also remained intact after controlling for the averaged
\(F(2,22) = 6.19; P < 0.01; \eta^2_p = 0.36\). Furthermore, the training-time-dependent gain in \(Gf\)
also remained intact after controlling for the averaged \(n\)-back
level reached in the last training session (ANCOVA: \(F(3,29) =
7.80; P < 0.001 \eta^2_p = 0.45\)).

In sum, these data indicate that the transfer effect on \(Gf\) scores
goes beyond an increase in working memory capacity alone. We
discuss this point in more detail below.

Discussion

We set the stage for our observed transfer effect by establishing
that there is an impressive learning curve for the training task in
all four experiments as expressed in comparable monotonically
increasing training functions across all of the training intervals.
These training results indicate that participants were challenged
and motivated to improve their performance even after a
training time as long as 4 weeks.

Having established a training effect, we then documented the
striking result of a training-related gain in \(Gf\), a finding that has
not been reported previously. How can such a transfer effect arise?

Operationally, we believe that the gain in \(Gf\) emerges because
of the inherent properties of the training task. The adaptive
character of the training leads to continual engagement of
executive processes while only minimally allowing the develop-
mant of automatic processes and task-specific strategies. As
such, it engages g-related processes (5, 17). Furthermore, the
particular working memory task we used, the ‘‘dual \(n\)-back’’ task,
engages multiple executive processes, including ones required to
inhibit irrelevant items, ones required to monitor ongoing per-
formance, ones required to manage two tasks simultaneously,
and ones required to update representations in memory. In
addition, it engages binding processes between the items (i.e.,
squares in spatial positions and consonants) and their temporal
context (30, 31).

Examining the transfer task in terms of the processes involved,
there is evidence that it shares some important features with the
training task, which might help to explain the transfer from the
training task to the \(Gf\) measures. First of all, it has been argued
that the strong relationship between working memory and \(Gf\)
primarily results from the involvement of attentional control
being essential for both skills (22). By this account, one reason
for having obtained transfer between working memory and
measures of \(Gf\) is that our training procedure may have facilit-
tated the ability to control attention. This ability would come
about because the constant updating of memory representations
with the presentation of each new stimulus requires the engage-
mant of mechanisms to shift attention. Also, our training task
discourages the development of simple task-specific strategies
that can proceed in the absence of controlled allocation of
attention.

Carpenter et al. (1) have proposed that the ability to abstract
relations and to maintain a large set of possible goals in working
memory accounts for individual differences in tasks such as the
Raven’s Advanced Progressive Matrices test, and therefore in
\(Gf\). This ability to maintain multiple goals in working memory
seems especially crucial in speeded \(Gf\) tasks because one can
speed performance by maintaining more goals in mind at once
to foster selection among representations. Therefore, after
training working memory, participants should be able to come
up with more correct solutions within the given time limit of our
speeded version of the \(Gf\) task.

However, our additional analyses show that there is more to
transfer than mere improvement in working memory capacity in
that the increase in \(Gf\) was not directly related to either
preexisting individual differences in working memory capacity
or to the gain in working memory capacity as measured by simple
or complex spans, or even, by the specific training effect itself.

Therefore, it seems that the training-related gain on \(Gf\) goes
beyond what sheer capacity measures even if working memory
capacity is relevant to both classes of tasks. Of course, tasks that
measure \(Gf\) are picking up other cognitive skills as well, and
perhaps the training is having an effect on these skills even if
measures of capacity are not sensitive to them. One example
might be multiple-task management skills. Our dual \(n\)-back task
requires the ability to manage two \(n\)-back tasks simultaneously,
and it may be this skill that is common to tasks that measure \(Gf\).
Our measures of working memory capacity, by contrast, index
capacity only for simpler working memory tasks that are not so
demanding of multiple-task management skills. So, sheer working memory capacity alone may be an important component of measures of Gf, but beyond this capacity, there may be other skills not measured by simpler working memory tasks that are engaged by the training task and that train skills needed in measures of Gf. It may still be the case that training on the dual n-back task promotes development of these non-capacity skills. A line of evidence consistent with this idea shows that individual differences in both working memory span and in n-back tasks are related to individual differences in Gf (23, 25, 32).

The finding that the transfer to Gf remained even after taking the specific training effect into account seems to be counterintuitive, especially because the specific training effect is also related to training time. The reason for this capacity might be that participants with a very high level of n at the end of the training period may have developed very task specific strategies, which obviously boosts n-back performance, but may prevent transfer because these strategies remain too task-specific (5, 20). The averaged n-back level in the last session is therefore not critical to predicting a gain in Gf; rather, it seems that working at the capacity limit promotes transfer to Gf.

Of particular importance is the finding that preexisting inter-individual differences in Gf as measured in the pretest are not related to the training-related gain in Gf. This finding indicates that the effect of training is not restricted to participants within a certain range of cognitive abilities. Both initial low-Gf as well as initial high-Gf participants profit from training similarly. Still, although the interaction was not reliable, we remain cautious about this result because numerically the low-Gf participants showed somewhat larger gains than the high-Gf participants. Of course, this result may be accounted for by regression to the mean, but it may also be that the training was of truly greater benefit to lower Gf participants, if not reliably so in our study.

The dose-responsive gain in Gf indicates that the training benefit is not a threshold phenomenon. The constraints of our experiments do not permit us to know how much longer we could have continued training before failing to realize any further gains in Gf. This dose responsiveness is an important issue for further study because the exact plot of gain with training could have important practical implications for those interested in training fluid intelligence. Our study also does not permit us to know how long the training gain persists; longitudinal studies will be required to address that issue.

These limitations notwithstanding, our findings are of general significance because they provide evidence for enhancement of fluid intelligence by cognitive training different from training the test itself. The finding that cognitive training can improve Gf is a landmark result because this form of intelligence has been claimed to be largely immutable. Instead of regarding Gf as an immutable trait, our data provide evidence that, with appropriate training, there is potential to improve Gf. Moreover, we provide evidence that the amount of Gf-gain critically depends on the amount of training time. Considering the fundamental importance of Gf in everyday life and its predictive power for a large variety of intellectual tasks and professional success, we believe that our findings may be highly relevant to applications in education.

Materials and Methods

Participants and Procedure. For this study, we conducted four individual experiments involving a total of 70 healthy young participants (36 female; mean age, 25.6 years of age; SD, 3.3) recruited from the University of Bern community, 35 of whom performed working-memory training in four different training settings (one of the participants failed to complete the required training sessions and was thus discarded from the data analysis, resulting in a final N of 34). These training groups were matched to four control groups who did not have training (n = 35). The crucial difference among the four training settings was the number of training sessions between pre- and posttests, ranging from 8 to 19 sessions (i.e., 8 days (n = 16), 12 days (n = 22), 17 days (n = 16), and 19 days (n = 15)), with the control groups receiving the pre- and posttesting at the same intervals as the trained groups. In each training setting, participants trained daily, except for the weekends. The posttest took place at least 1 day after the last training session, with the largest interval being 2 days.

Materials. Training task. For the training task, we used the same material as described by Jaeggi et al. (33), which was a dual n-back task where squares at eight different locations were presented sequentially on a computer screen at a rate of 3 s (stimulus length, 500 ms; interstimulus interval, 2,500 ms). Simultaneously with the presentation of the squares, one of eight consonants was presented sequentially through headphones. A response was required whenever one of the presented stimuli matched the one presented n positions back in the sequence. The value of n was the same for both streams of stimuli. There were six auditory and six visual targets per block (four appearing in only one modality, and two appearing in both modalities simultaneously), and their positions were determined randomly. Participants made responses manually by pressing on the letter “A” of a standard keyboard with their left index finger for visual targets, and on the letter “L” with their right index finger for auditory targets. No responses were required for non-targets.

In this task, the level of difficulty was varied by changing the level of n (34), which we used to track the participants’ performance. After each block, the participants’ individual performance was analyzed, and in the following block, the level of n was adapted accordingly: If the participant made fewer than three mistakes per modality, the level of n increased by 1. It was decreased by 1 if more than five mistakes were made, and in all other cases, n remained unchanged.

A training session comprised 20 blocks consisting of 20 + n trials resulting in a daily training time of ~25 min.

Transfer tasks. We used standardized fluid intelligence tests, consisting of visual analogy problems of increasing difficulty. Each problem presents a matrix of patterns in which one pattern is missing. The task is to select the missing pattern among a set of given response alternatives. For the experiment with eight training sessions, we used the Raven’s Advanced Progressive Matrices (RAPM) test, set II (35), whereas for all other experiments, we used the short version of the Bochumer Matrizen-Test (BOMAT) (36), a more difficult variant of the RAPM. For the RAPM, we used parallel forms for the pre- and posttesting by dividing the test into even and odd items (24); for the BOMAT, we used the published A and B versions. To keep the pre- and posttest sessions short enough, we allowed limited time (10 min) to complete the task, and the number of correct solutions provided in that time served as the dependent variable.

To control for the impact of individual differences and gain in working memory capacity, a digit-span task (38), as well as a reading span task (39), was used in the pre- and postsession. However, the reading span task was not assessed in the 8-day group.

5. te Nijenhuis J, van Vianen AEM, van der Flier H (2007) Score gains on g-loaded tests: No g. Intelligence 35:283–300.

ACKNOWLEDGMENTS. We thank Daniela Blaser and Oliver Markes for help with data collection and the participants for their time and effort. The preparation of this manuscript was supported by Swiss National Science Foundation Fellowships PA001-117473 (to S.M.J.) and PBEB1-117527 (to M.B.) and by grants from the National Science Foundation and the National Institute of Mental Health (to J.J.).

6832 | www.pnas.org/cgi/doi/10.1073/pnas.0801268105
Jaeggi et al.