Academic and social-emotional learning

By Maurice J. Elias

EDUCATIONAL PRACTICES SERIES–11

UNESCO

The International Academy of Education - IAE

The International Academy of Education (IAE) is a not-for-profit scientific association that promotes educational research, its dissemination, and the implementation of its implications. Founded in 1986, the Academy is dedicated to strengthening the contributions of research, solving critical educational problems throughout the world, and providing better communication among policy-makers, researchers and practitioners. The seat of the Academy is at the Royal Academy of Science, Literature and Arts in Brussels, Belgium, and its co-ordinating centre is at Curtin University of Technology in Perth, Australia.

The general aim of the IAE is to foster scholarly excellence in all fields of education. Towards this end, the Academy provides timely syntheses of research-based evidence of international importance. The Academy also provides critiques of research, its evidentiary basis, and its application to policy.

The current members of the Board of Directors of the Academy are:

- Erik De Corte, University of Leuven, Belgium (*President*)

- Herbert Walberg, University of Illinois at Chicago, United States of America (*Vice President*)

- Barry Fraser, Curtin University of Technology, Australia (*Executive Director*)

- Jacques Hallak, Paris, France

- Michael Kirst, Stanford University, United States of America

- Ulrich Teichler, University of Kassel, Germany

http://www.curtin.edu.au/curtin/dept/smec/iae

Series preface

This booklet is about the social-emotional skills students need for success in school and in life. In any classroom in the world, from the simplest, with no walls, to the most elaborate, teachers must get along with students and students must get along with one another if learning is to take place. Social-emotional skills, or 'emotional intelligence', is the name given to the set of abilities that allows students to work with others, learn effectively, and serve essential roles in their families, communities and places of work.

Research shows that social-emotional skills can be taught to students and that their presence in classrooms and schools improves academic learning. When academic and social-emotional learning both become a part of schooling, students are more likely to remember and use what they are taught. They also incorporate into their education a sense of responsibility, caring, and concern for the well being of others, as well as themselves. Learning thus can be said to touch both the 'head' and the 'heart' and the result is classrooms that are run better and students who are more inspired. Academic and social-emotional learning are therefore connected in every school, worldwide.

Much also has been learned about how to enhance academic and social-emotional learning in ways that are more likely to work well. This booklet gives the principles that have been shown to lead to success. It contains important guidelines for building academic and social-emotional skills, and sections in each chapter on practical applications that can be brought into classrooms and schools without difficulty. In addition, there is an extensive section on resources, including international resources that are accessible via the Internet.

This booklet has been prepared for inclusion in the Educational Practices Series developed by the International Academy of Education and distributed by the International Bureau of Education and the Academy. As part of its mission, the Academy provides timely syntheses of research on educational topics of international importance. This booklet is the eleventh in the series on educational practices that generally improve learning.

The author of this booklet is Maurice J. Elias, a Professor of Psychology at Rutgers University and Vice-Chair of the Leadership Team of the Collaborative for Academic, Social, and Emotional Learning (www.CASEL.org). With colleagues at CASEL, Elias was senior author of *Promoting social and emotional learning: guidelines for educators*, published by the Association for Supervision and Curriculum Development and circulated to over 100,000 educational leaders internationally. Author of over a dozen books and numerous

articles and book chapters, Elias has also written for newspapers and magazines, including an award-winning column for parents. Elias' books have been translated into over a dozen languages and he has lectured in Asia, Europe, the Middle East, and throughout North America.

The following individuals served as reviewers and suggested improvements in a draft version of this booklet to make it more applicable to various cultures, and easier to understand and translate: Milton Chen is Executive Director of the George Lucas Educational Foundation, San Rafael, CA, which creates media and a web site (www.glef.org) to promote success stories in education and the use of technology; Mario Luis Pacheco Filella is manager of technology and educational development at the Santa Engracia, Mexico, Hospital, Health Division, Pulsar Group and is affiliated with Duxx, the Graduate School of Business Leadership in Monterrey; Keisha Mitchell, a Jamaican national, is a Ph.D. candidate in psychology at Rutgers University in New Brunswick, New Jersey. Her interests include social-emotional learning and social support, and education as an agent for community change; Kathariya Mokrue was born in Bangkok, Thailand. She is presently a pre-doctoral intern at Montefiore Medical Centre in New York City. Her research examines social-emotional education, family support and coping; Leslie Swartz, Ph.D., is director, child/youth and family development at the Human Sciences Research Council, South Africa, and professor of psychology at the University of Stellenbosch.

The officers of the International Academy of Education are aware that this booklet is based on research carried out primarily in economically advanced countries. However, the booklet focuses on aspects of learning that are universal. The reviewers noted that those who live in countries with a stable educational infrastructure (i.e. uninterrupted schooling, adequate access and materials, and children whose families are not in deep poverty or suffering from epidemics) will be able to carry out all of the suggestions, but, as one put it, 'I am certain that this booklet—as it is—can widen horizons and be applied to create awareness and practice of social-emotional intelligence that is so much needed, in all of our nations.' The practices presented here are likely to be generally applicable throughout the world. Even so, the principles should be assessed with reference to local conditions, and adapted accordingly. In any educational setting or cultural context, suggestions or guidelines for practice require sensitive and sensible application, and continuing evaluation.

HERBERT J. WALBERG
Editor, IAE Educational Practices Series
University of Illinois at Chicago

Table of contents

Introduction

In every society, children will inherit social roles now occupied by adults. Our education systems have the job of preparing children for this eventual responsibility. Therefore, around the world, people want to improve education. Some want to strengthen basic academic skills; others want to focus on critical thinking. Some want to promote citizenship or character; others want to protect children against the dangers of drugs, violence and alcohol. Some want parents to play a larger role; others feel the entire community should be involved.

There are some areas of growing consensus. As indicated by numerous polls of parents and community leaders, we are clear what we want our children to know and to be able to do, and this defines what we want schools to teach. We want young people to:

- Be fully literate, able to benefit from and make use of the power of written and spoken language, in various forms;
- Understand mathematics and science at levels that will prepare them for the world of the future and strengthen their ability to think critically, carefully and creatively;
- Be good problem-solvers;
- Take responsibility for their personal health and well-being;
- Develop effective social relationships such as learning how to work in a group and how to understand and relate to others from different cultures and backgrounds;
- Be caring individuals with concern and respect for others;
- Understand how their society works and be prepared to take on the roles that are necessary for future progress;
- Develop good character and make sound moral decisions.

All of these are aspects of what some refer to as the 'education of the whole child'. Educating the whole child is not a new idea. It is rooted in the writings and teachings of many ancient cultures. Yet, achieving the kind of balance that encourages all children to learn, work and contribute to their fullest potential has been a continuing challenge as our world has grown more complex and our communities more fragmented. The final six points refer to aspects of education that have been referred to as character education, service learning, citizenship education and emotional intelligence. All of these can be expressed in the single term, social-emotional learning, and it is this form of education, when added to academic learning, that provides

educators with the possibility of capturing the balance children need.

While some may disagree about what is most important, educators, parents, business leaders and those who make social policy share the same set of concerns. Schools must become better at guiding children toward becoming literate, responsible, non-violent, drug-free and caring adults.

The challenge of raising literate, responsible, non-violent, drug-free and caring children is familiar to parents, policy makers, administrators and teachers. Experience and research show that each element of this challenge can be enhanced by thoughtful, sustained and systematic attention to the social and emotional skills of children. Indeed, schools worldwide must give children intellectual and practical tools they can bring to their classrooms, families and communities. Social-emotional learning provides many of these tools. It is a way of teaching and organizing classrooms and schools that help children learn a set of skills needed to manage life tasks successfully, such as learning, forming relationships, communicating effectively, being sensitive to others' needs and getting along with others. When schools implement high-quality social-emotional learning programmes effectively, the academic achievement of children increases, incidences of problem behaviours decrease, and the relationships that surround each child are improved.

Social-emotional learning is sometimes called 'the missing piece', because it represents a part of education that links academic knowledge with a specific set of skills important to success in schools, families, communities, workplaces and life in general. As recent world events have taught, there is a danger to each of us—locally and globally—when children grow up with knowledge but without social-emotional skills and a strong moral compass. Hence, a combination of academic and social-emotional learning is the true standard for effective education in the world today and for the foreseeable future.

1. Learning requires caring

Effective, lasting academic and social-emotional learning is built upon caring relationships and warm but challenging classroom and school environments.

Research findings

Lasting social-emotional learning, sound character and academic success are founded on classrooms and schools that are not threatening to students and challenge them to learn more, but do so in ways that do not discourage them. Also, these schools are places where students feel cared about, welcomed, valued and seen as more than just learners—they are seen as resources.

Pratical applications

- Greet all students by name when they enter the school or classroom.
- Begin and/or end the school day with brief periods of time for students to reflect on what they have learned recently and what they might want to learn next.
- Create rules in the classroom that recognize positive behaviour, such as co-operation, caring, helping, encouragement and support. Be sure that discipline rules and procedures are clear, firm, fair and consistent.
- Show interest in their personal lives outside the school.
- Ask them what kinds of learning environments have been most and least successful for them in the past and use this information to guide instruction.

Suggested readings: Kriete & Bechtel, 2002; Lewis, Schaps & Watson, 1996; O'Neil, 1997; Osterman, 2000.

2. Teach everyday life-skills

Life-skills that promote academic and social-emotional learning must be taught explicitly in every grade level.

Research findings

The Collaborative for Academic, Social, and Emotional Learning (www.CASEL.org) has identified a set of social-emotional skills that underlie effective performance of a wide range of social roles and life tasks. To do this, CASEL drew from extensive research in a wide range of areas, including brain functioning, and methods of learning and instruction. These are the skills that provide young people with broad guidance and direction for their actions in all aspects of their lives, in and out of school. The skills are included below.

CASEL's essential skills for academic and social-emotional learning

Know yourself and others:
- Identify feelings—recognize and label one's feelings;
- Be responsible—understand one's obligation to engage in ethical, safe and legal behaviours;
- Recognize strengths—identify and cultivate one's positive qualities.

Make responsible decisions:
- Manage emotions—regulate feelings so that they aid rather than impede the handling of situations;
- Understand situations—accurately understand the circumstances one is in;
- Set goals and plans—establish and work toward the achievement of specific short- and long-term outcomes;
- Solve problems creatively—engage in a creative, disciplined process of exploring alternative possibilities that leads to responsible, goal-directed action, including overcoming obstacles to plans.

Care for others:
- Show empathy—identifying and understanding the thoughts and feelings of others;

- Respect others—believing that others deserve to be treated with kindness and compassion as part of our shared humanity;
- Appreciate diversity—understanding that individual and group differences complement one another and add strength and adaptability to the world around us.

Know how to act:
- Communicate effectively—using verbal and non-verbal skills to express oneself and promote effective exchanges with others;
- Build relationships—establishing and maintaining healthy and rewarding connections with individual and groups;
- Negotiate fairly—achieving mutually satisfactory resolutions to conflict by addressing the needs of all concerned;
- Refuse provocations—conveying and following through effectively with one's decision not to engage in unwanted, unsafe, unethical behaviour;
- Seek help—identifying the need for and accessing appropriate assistance and support in pursuit of needs and goals;
- Act ethically—guide decisions and actions by a set of principles or standards derived from recognized legal/professional codes or moral or faith-based systems of conduct.

Practical applications

- Consider adopting a social-emotional skill-building programme that has shown demonstrated effectiveness in populations and circumstances similar to yours; listings and Internet links to listings are available at www.CASEL.org, www.NASPonline.org and in the 'Resources' section of this booklet.
- Use CASEL's list of skills to help students prepare for academic assignments, projects, homework and tests.
- Ask students when it is important in their lives to use each of the skills. Then, help them build and use the skills when these situations arise.
- Each week, try to incorporate building one skill on CASEL's list of skills into your usual instructional routine. Continue throughout the year, reviewing and deepening what you do as you repeat each skill.

Suggested readings: Connell et al., 1986; Elias et al., 1997; Elias, Tobias & Friedlander., 2000; Goleman, 1995; Topping & Bremner, 1998; Zins et al., 2003.

3. Link social-emotional instruction to other school services

Application of social-emotional skills to everyday life is aided greatly by a consistent, developmentally appropriate structure of supportive services in the school.

Research findings

In addition to teaching life-skills explicitly at elementary and secondary levels, children also benefit from co-ordinated, explicit, developmentally sensitive instruction in the prevention of specific problems, such as smoking, drug use, alcohol, pregnancy, violence and bullying. Different cultures will select and focus on preventing different problem behaviours. In a similar way, children benefit from explicit guidance in finding a healthy life style. Eating habits, sleeping patterns, study and work environments are among the areas that are important to promoting academic and social-emotional learning. Further, all students need to be taught and given opportunities to practice age-appropriate strategies for conflict resolution. Finally, schools should be attentive to difficult life events that befall students and try to provide them with support and coping strategies at those stressful moments. Typically, such assistance is not given until children show problems that are the result of those difficult life events; unfortunately, during this time, many students are distracted from learning. Even when they are not actively disrupting class, they are not taking in all that their teachers are working so hard to provide. Providing social-emotional assistance to children facing difficult events is a sound prevention strategy that also promotes better academic learning. Children with special education needs must also receive social-emotional skill-building instruction and be included in related activities.

Practical applications

- Provide time in the school curriculum each year for instruction in appropriate health issues and problem behaviour prevention.

- Organize guidance and counselling services so that they help build social-emotional skills of groups of children who are anticipating or facing difficult situations.
- Allow planning time for staff to co-ordinate their efforts at supporting academic and social-emotional learning.

Suggested readings: Adelman & Taylor, 2000; Comer et al., 1999; Elias et al., 1997; Jessor, 1993; Perry & Jessor, 1985.

4.   Use goal-setting to focus instruction

> Goal-setting and problem-solving provide direction and energy for learning.

Research findings

Children are required to learn many things, but without a sense of connection between and to those things, children are not likely to retain what they learn and use it in their lives. When their learning is presented in terms of understandable goals (goals that children can play a larger role in defining as they get older), children become more engaged and focused and less likely to exhibit behaviour problems.   Learning experiences that co-ordinate and integrate different aspects of learning across subject areas and over time, as well as those that link to their lives outside of school in the present and future, are especially valuable.

Children also benefit from learning problem-solving strategies that they can apply to new situations that face them. Instruction in reading that includes examining the problem-solving and decision-making processes used by various characters in stories is particularly enriching.   The same is true for history and current events instruction that allows students to focus on the different perspectives of individuals and groups involved and the problem-solving processes they used (or might have used). A similar approach can be used to help students understand how scientific and mathematical problem-solving occurs. When taught in this way, students find that, as they encounter new books, new civic situations and new group processes, they have strategies to apply that enhance their learning and performance and enable them to make better progress.

Practical applications

- Ask students how they calm themselves down when they are very upset; remind them to use this strategy when they get into frustrating or difficult situations, or teach them a self-calming strategy.

- Have students set goals that include how they will get better at a particular area of study or schooling and how they will make a contribution to the classroom.
- Teach a problem-solving strategy for understanding fiction, history or current events that uses frameworks such as those illustrated in the following examples or related ones.

Here is an example that can be used for history. It can easily be adapted for discussion of current events.

Thinking about important events in history

- What is the event that you are thinking about? When and where did it happen? Put the event into words as a problem or choice or decision.
- What people or groups were involved in the problem? What were their different feelings? What were their points of view about the problem?
- What did each of these people or groups want to have happen? Try to put their goals into words.
- For each person or group, name some different options or solutions to the problem that they thought might help them reach their goals.
- For each option or solution, picture all the things that might have happened next. Envision both long- and short-term consequences.
- What were the final decisions? How were they made? By whom? Why? Do you agree or disagree? Why?
- How was the solution carried out? What was the plan? What obstacles or roadblocks were met? How well was the problem solved? Why?
- Rethink it. What would you have chosen to do? Why?

Here is an example that can be used for reading stories in elementary school. It can be combined with elements of the history framework to be more challenging as students get older.

- I will write about this character ...
- My character's problem is ...
- How did your character get into this problem?
- How does the character feel?
- What does the character want to happen?
- What are all the ways the character can get this to happen?
- What questions would you like to be able to ask to the character you picked, to one of the other characters, or to the author?

Suggested readings: Cohen, 1999; Elias et al., 1997; Elias & Tobias, 1996; Pasi, 2001; Topping & Bremner, 1998.

5. Use varied instructional procedures

> Instruction for academic and social-emotional learning should use varied modalities and approaches to reach the diverse styles and preferences of all learners.

Research findings

Academic and social-emotional learning takes place best in different ways for different students. So, educational experiences marked by instruction that uses different modalities are most likely to reach all children and allow them to build their skills and feel that the classroom environment is suited to their preferred way of learning. Modalities include modelling, role-playing, art, dance, drama, working with materials and manipulatives, and digital media, computer technology, and the Internet. Also important for sound instruction are regular and constructive feedback, discussions that include open-ended questioning, and frequent reminders to use social-emotional skills in all aspects of school life.

Practical applications

- Use a balance of teaching strategies, including asking open-ended questions, suggesting possible answers from which students might choose, checking with students to see if they understand what has been taught by asking them to repeat it to you or to a classmate, role-playing and lecturing.
- Vary instruction so that sometimes students are working in a large group, in small groups, in pairs, by themselves, at the computer, or on the Internet, working with digital media.
- Provide opportunities for cross-age tutoring.
- Create learning centres so that students can move around and have different learning experiences over the course of a day. The centres can be related to Howard Gardner's concept of multiple intelligences, so that some can be very tactile and hands-on, others can involve writing, others can

relate to art or music, and others can provide opportunities to use to dramatic or imaginative play.

- Allow students to create exhibitions of what they learn in different subject areas that can be shared with other students, parents and members of the community.
- Bring in experts and other individuals in the community to share knowledge, skills, customs and stories with students.

Suggested readings: Gardner, 2000; Johnson & Johnson, 1994; Ladd & Mize, 1983; Lambert & McCombs, 1998; Noddings, 1992; Salovey & Sluyter, 1997; Topping, 2000.

6. Promote community service to build empathy

Community service plays an essential role in fostering generalization of social-emotional skills, particularly in building empathy.

Research findings

Properly conducted community service, which begins at the earliest level of schooling and continues throughout all subsequent years, provides an opportunity for children to learn life-skills, integrate them, apply them, reflect upon them and then demonstrate them. This process solidifies their learning and also helps to create a climate in which others are more likely to engage in community service. Service experiences usually help students to encounter other people, ideas and circumstances in ways that broaden their sense of perspective and build empathic understanding and caring connections to the world around them. For many young people, community service provides an opportunity to nourish a universal need to be a generous and contributing member of important groups to which one belongs. This helps prepare children for their eventual roles in the larger society, as well as work and family groups of which they will be a part. Further, it helps nurture the spirit of students to see themselves as part of a larger world, with sets of ideals and beliefs that are important to living a fulfilled life.

Practical applications

- Provide service opportunities within classrooms so that, even from the youngest age, students feel that they are making a contribution to the positive functioning of the classroom. Examples include putting chairs away, cleaning up, and helping the teacher and other students.
- Set up opportunities for students to take on helpful roles in the community. Examples include improving the physical environment around the school, helping the elderly, and

providing comfort and support to the injured or sick. Such opportunities begin with preparation, so that students understand the circumstances they will be involved with, for example, the kinds of illnesses and difficulties that beset the elderly. Then, there is the action of carrying out the service, in which students should be as directly involved as is appropriate to their age and safety. Action is followed by reflection, as students have a chance to talk and/or write about what they experienced and their feelings about it. Finally, demonstration of learning should take place, as students creatively show their peers, younger students, parents and/or other groups in the community what they did, why they did it, how they felt about it and what they learned.

Suggested readings: Berman, 1997; Billig, 2000; National Commission on Service Learning, 2002.

7.   Involve parents

Involvement of parents in partnerships with the school to promote students' academic and social-emotional learning is likely to improve results.

Research findings

When home and school collaborate closely to implement social-emotional learning programmes, students gain more and programme effects are more enduring and pervasive. As more and more children are being bombarded by messages of mass culture, the Internet, television, music, videos and other outlets unfiltered by adults, it becomes more and more important that key caregivers in children's lives send strong and co-ordinated health-promoting messages. Parents, schools, the community and the larger society all agree that building children's social-emotional skills is an important common concern. In the resources section of this document, there is a book to help parents create a positive home climate, build children's social-emotional skills, and organize so that family responsibilities and schoolwork all get accomplished.   Entitled *Emotionally intelligent parenting*, it is available in at least ten different languages worldwide as of the time of this writing (November 2002).

Practical applications

- Give parents regular overviews of the academic and social-emotional skills students are learning at any given time.
- Give parents opportunities to meet to exchange ideas about how to support the teaching in school and how to raise their children.
- Help parents learn how to organize the morning routine and homework routines to minimize conflict.
- Communicate to parents the importance of having positive times with their children, despite difficulties, in order to build the children's sense of hope.
- Provide parents with opportunities to contribute to the classroom and/or school on a regular basis.

- Create a welcoming climate for parents in the school by displaying student artwork and other projects near entranceways.
- Set up time for family instruction or family projects, when parents and students can work together in appropriate ways.

Suggested readings: Christenson & Havsy, 2003; Elias, Tobias & Friedlander, 2000; Epstein, 2001; Huang & Gibbs, 1992.

8. Build social-emotional skills gradually and systematically

Implementation of social-emotional learning into a school is an innovation that should be built on the existing strengths of the setting and occurs in stages over a period of several years.

Research findings

Selecting and implementing social-emotional learning programmes should follow after a consideration of local needs, goals, interests and mandates; staff skills, workload and receptiveness; pre-existing instructional efforts and activities; the content and quality of programme materials; its developmental and cultural appropriateness to the range of recipient student populations; and its acceptability to parents and community members. Social-emotional learning efforts are often implemented as pilot projects and it typically takes two or three years for staff to have a confident and competent sense of ownership of the approaches being used. Once implemented, these efforts are most likely to become a regular part of school schedules and routines to the extent to which they are aligned with local and national educational goals, comply with legal standards and mandates, and have the informed support of educational administration, organized groups of educators, and members of the community or government who oversee high-quality education. Of particular importance is the connection between academic and social-emotional learning. Social-emotional learning is not a separate subject area; rather, it must be linked to language literacy, instruction in mathematics and science, history and current culture, health and physical education, and the performing arts. In all of these areas, the essential skills for academic and social-emotional learning mentioned earlier allow for deeper understanding of the content and improved pedagogy, with greater student engagement in learning and fewer behaviour disruptions.

Teachers and parents often complain that students do not adopt the goals they hold for them, and that they do not follow up on their well-meant advice. For example, Stefano's father tries to prevent him from doing his homework with the radio on, believing that music affects motivation and performance negatively. Current research does not support this view. Yet, such conflicts of interest lead to the frustration of Stefano's need for autonomy. Often, teachers (and parents) try to push their own goals along, thus fueling the child's struggle for autonomy. For decades, schools, teachers and researchers narrowed educational goals to learning and achievement, which only frustrated students' social goals.

Practical applications

- Allocate time and resources to those who are involved in programme planning, co-ordination and leadership.
- Develop a policy that states clearly how academic and social-emotional learning fit together in the schools.
- Begin social-emotional learning efforts with small, pilot projects conducted by those best trained in principles of social-emotional instruction and programmes.
- Allow time to work with the results of pilot projects to plan expanded efforts and/or new pilot projects.

Suggested readings: Collaborative for Academic, Social, and Emotional Learning, 2002; Elias et al., 1997; Novick, Kress & Elias, 2002; Utne O'Brien, Weissberg & Shriver, 2003.

9.   Prepare and support staff well

> Effective academic and social-emotional instruction follows from well-planned professional development for all school personnel and a system of support during the initial years of implementation.

Research findings

Social-emotional learning is relatively new to many educators. Therefore, they need to be patient with themselves and allow themselves an opportunity to learn this new area. No lasting success in academic and social-emotional instruction can be expected without on-going professional development for school personnel and support for their efforts as implementation proceeds. Time should be taken to train staff in children's social-emotional development, modelling and practice of effective teaching methods, multi-modal instruction, regular coaching and constructive feedback from colleagues. Staff also should become familiar with best practices in the field so that teachers can draw on what works most effectively. CASEL is playing a significant role in identifying the best of what works. Its guide, *Safe and sound*, is available on the Internet and provides guidelines and information to allow educators to find programmes and procedures that work best for their particular situations.

Practical applications

- Provide high-quality staff development and support in social-emotional programmes and instructional procedures for those carrying out social-emotional learning efforts.
- Provide related professional development for all school personnel, including training in how to develop school-wide efforts to promote social-emotional learning skills.
- Create a committee that will be responsible for supporting implementation, especially during the initial years.

Suggested readings: Collaborative for Academic, Social, and Emotional Learning, 2002; Kessler,

10.   Evaluate what you do

Evaluation of efforts to promote social-emotional learning is an ethical responsibility that involves on-going monitoring of implementation, assessing outcomes, and understanding opinions and reactions of those who carry out and receive the efforts.

Research findings

When schools accept children through their doors, they are making a pledge to prepare those students for the future. While schools cannot guarantee the outcomes of all their efforts, they do have an ethical responsibility to monitor what they do and to attempt to continuously improve it. Therefore, schools need ways to keep track of student learning and performance in all areas, including the development of social-emotional abilities. Socio-emotional learning efforts should be monitored regularly, using multiple indicators to ensure programmes are carried out as planned. In addition, on-going programme outcome information and consumer satisfaction measures can be systematically gathered from multiple sources.   Instruction must be adapted to changing circumstances. This occurs through examining the opinions of those delivering and receiving social-emotional instruction; documenting ways in which social-emotional programmes are implemented and connected with academic instruction; evaluating outcomes observed among various groups of children in one's schools; and monitoring and addressing on-going new developments, such as changes in district resources, state initiatives and scientific advances.

Practical applications

- Use checklists to keep track of whether socio-emotional learning activities that are planned actually take place.
- Provide staff with the opportunity to rate and/or comment on the lessons they carry out, to note what went well and what might be improved in the future.

- Use brief surveys of students and staff to find out what they liked most and least about socio-emotional learning activities, the opportunities they had for putting the skills to use, and ideas for improving instruction.
- Ask people who work in the school (and parents, if possible) how they will know when students' academic and social-emotional skills are improving and design indicators to measure the extent to which this takes place.
- Place on the report card or other feedback system a listing of socio-emotional learning skills or related indicators so that there can be accountability for this aspect of schooling and methods designed to improve instruction as needed.

Suggested readings: Elias et al., 1997; Fetterman, Kafterian & Wandersman, 1996; Harvard Graduate School of Education, 2003; Weissberg et al., 1997.

Conclusion

Education is changing. Academic and social-emotional learning is becoming the new standard for what is considered the basics that children should acquire during their schooling. Because this is so new to many educators, but not to all, this pamphlet includes ideas to help get social-emotional efforts started, as well as to sustain those schemes that have already begun. It is designed to help all schools become places where learning is valued, dreams are born, leaders are made, and the talents of students—the greatest resource shared by every community—are unleashed.

Our students are important not only to their schools and families, but also to their communities, their future workplaces and families, and to the world around them. Each student has potential. While that potential is not identical for all, every student deserves the opportunity to have his potential developed. The combination of academic and social-emotional learning is the most promising way to accomplish this goal. In so doing, educators are also preparing students for the tests of life, for the responsibilities of citizenship, and for adopting a lifestyle that is literate, responsible, non-violent, drug-free, and caring. This is not an easy task. It will require patience as new skills are learned, but not too much patience, as our students are depending on the adults around them to prepare them for their future lives. It is a great responsibility, and it deserves great effort.

References

Adelman, H.S.; Taylor, L. 2000. Moving prevention from the fringes into the fabric of school improvement. *Journal of education and psychological consultation* (Mahwah, NJ), vol. 11, no. 1, p. 7–36.

Berman, S. 1997. *Children's social consciousness and the development of social responsibility*. Albany, NY, State University of New York. (SUNY series: Democracy and education).

Billig, S. 2000. The impact of service learning on youth, schools, and communities: research on K-12 school-based service learning, 1990-1999. Available from:http://www.learningindeed.org/research/slreseaerch/slrschsy.html

Christenson, C.L.; Havsy, L.H. 2003. Family-school-peer relationships: significance for social, emotional, and academic learning. *In:* Zins, J.E, et al., eds. *Building school success on social and emotional learning*. New York, NY, Teachers College Press.

Cohen, J., ed. 1999. *Educating minds and hearts: social emotional learning and the passage into adolescence*. New York, NY, Teachers College Press.

Collaborative for Academic, Social, and Emotional Learning. 2002. *Safe and sound: An an educational leader's guide to evidence-based social and emotional learning programs*. Chicago, IL, AuthorCASAL.

Comer, J. P., et al., eds. 1999. *Child by child: the Comer process for change in education*. New York, NY, Teachers College Press.

Connell, D.B., et al. 1986. School health education evaluation. *International journal of educational research* (Tarrytown, NY), vol. 10, p. 245–345.

Elias, M.J.; Tobias, S.E. 1996. *Social problem solving interventions in the schools: curriculum materials for educators*. Distributed by National Professional Resources: www.nprinc.com.

Elias, M.J.; Tobias, S.E.; Friedlander, B.S. 2000. *Emotionally intelligent parenting: how to raise a self-disciplined, responsible, socially skilled child*. New York, NY, Random House/Three Rivers Press.

Elias, M.J., et al. 1997. *Promoting social and emotional learning: guidelines for educators*. Alexandria, VA, Association for Supervision and Curriculum Development. [Also in Japanese.]

Epstein, J.L. 2001. *School, family, and community partnerships: preparing educators and improving schools*. Boulder, CO, Westview Press.

Fetterman, D.M.; Kaftarian, S. J.; Wandersman, A. 1996. *Empowerment evaluation: knowledge and tools for self-assessment and accountability*. Newbury Park, CA, Sage.

Gardner, H. 2000. *Intelligence reframed: multiple intelligences for the 21st century*. New York, NY, Basic Books.

Goleman, D. 1995. *Emotional intelligence: why it can matter more than IQ*. New York, NY, Bantam Books. [Available in many international editions.]

Harvard Graduate School of Education. 2003. *The evaluation exchange*. Cambridge, MA:www.gse.harvard.edu/hfrp/eval/archives.html.

Huang, L.; Gibbs, J. 1992. Partners or adversaries? Home-school collaboration across culture, race, and ethnicity. *In*: Christenson, S; Close Conoley, J., eds. *Home-school collaboration: enhancing children's academic and social competence*, p. 81–110. Silver Spring, MD, National Association of School Psychologists.

Jessor, R. 1993. Successful adolescent development among youth in high-risk settings. *American psychologist* (Washington, DC), vol. 48, p. 177–216.

Johnson, D.W.; Johnson, R.T. 1994. *Learning together and alone: cooperative, competitive, and individualistic learning*. Needham Heights, MA, Allyn & Bacon.

Kessler, R. 2000. *The soul of education: helping students find connection, compassion, and character at school*. Alexandria, VA, Association for Supervision and Curriculum Development.

Kriete, R.; Bechtel, L. 2002. *The morning meeting book*. Greenfield, MA, Northeast Foundation for Children.

Ladd, G.W.; Mize, J. 1983. A cognitive social-learning model of social-skill training. *Psychological review* (Washington, DC), vol. 90, p. 127–57.

Lambert, N.M.; McCombs, B.L., eds. 1998. *How students learn: reforming schools through learner-centered education.* Washington, DC, American Psychological Association.

Lantieri, L., ed. 2001. *Schools with spirit: nurturing the inner lives of children and teachers.* Boston, MA, Beacon Press.

Leiberman, A. 1995. Practices that support teacher development. *Phi delta kappan* (Bloomington, IN), vol. 76, p. 591–96.

Lewis, C.C.; Schaps, E.; Watson, M.S. 1996. The caring classroom's academic edge. *Educational leadership* (Alexandria, VA), vol. 54, p. 16–21.

National Commission on Service Learning. 2002. *The power of service learning.* Newton, MA, NCSL.

Noddings, N. 1992. *The challenge to care in schools: an alternative approach to education.* New York, NY, Teachers College Press.

Novick, B.; Kress, J.; Elias, M.J. 2002. *Building learning communities with character: how to integrate academic, social, and emotional learning.* Alexandria, VA, ASCD.

O'Neil, J. 1997. Building schools as communities: a conversation with James Comer. *Educational leadership* (Alexandria, VA), vol. 54, p. 6–10.

Osterman, K.F. 2000. Students' need for belonging in the school community. *Review of educational research* (Washington, DC), vol. 70, p. 323–67.

Pasi, R. 2001. *Higher expectations: promoting social emotional learning and academic achievement in your school.* New York, NY, Teachers College Press.

Perry, C.L.; Jessor, R. 1985. The concept of health promotion and the prevention of adolescent drug abuse. *Health education quarterly* (London, UK), vol. 12, p. 169–84.

Salovey, P.; Sluyter, D., eds. 1997. *Emotional development and emotional intelligence: educational implications.* New York, NY, Basic Books.

Topping, K. 2000. *Tutoring.* Geneva, Switzerland, International Bureau of Education and the International Academy of Education. [Educational practices series, booklet no. 5, see: www.ibe.unesco.org]

Topping, K.J.; Bremner, W.G. 1998. *Promoting social competence: practice and resources guide.* Edinburgh, UK, Scottish Office Education and Industry Department.

Utne O'Brien, M.; Weissberg, R.P.; Shriver, T.P. 2003. Educational leadership for academic, social, and emotional learning. *In:* Elias, M.J.; Arnold, H.; Hussey, C., eds. *EQ + IQ = Best leadership practices for caring and successful schools.* Thousand Oaks, CA, Corwin Press.

Weissberg, R.P.; et al., eds. 1997. *Healthy children 2010: establishing preventive services: issues in children's and families' lives,* vol. 9. Thousand Oaks, CA, Sage Publications.

Zins, J.E., et al., eds. 2003. *Building school success on social and emotional learning.* New York, NY, Teachers College Press.

Resources

This section contains web listings of international resources, materials, research, articles and training related to social-emotional learning.

GENERAL INTERNATIONAL RESOURCES

- Booklets in the series of Educational Practices published by the International Bureau of Education (translated into various foreign languages): www.ibe.unesco.org/International/Publications/EducationalPractices/prachome.htm
- The World Federation for Mental Health: www.wfmh.com/
- 6 Seconds: www.6seconds.org and www.heartskills.com/org/

PARTIAL LISTING OF NATIONAL WEB SITES

DENMARK
Center for Social and Emotional Learning (CESEL): www.cesel.dk/

GERMANY
German Network for Mental Health: www.gnmh.de/

ISRAEL
Psychological and Counseling Services/Life Skills Program (SHEFI), Education Ministry
http://www.education.gov.il/shefi

JAPAN
EQ Japan: www.eqj.co.jp/

NEW ZEALAND
Cornerstone Values: cornerstonevalues.org/
Youth Education Service (YES) of the New Zealand Police: www.nobully.org.nz/

NORWAY
Nasjonalforeningen for Folkehelsen:
www.nasjonalforeningen.no/BarnogFamilie/artikler/folkeskikk.htm

ROMANIA
Right Training and Consulting: right.netfirms.com/

SINGAPORE
Mega Forte Centre (The Centre for Emotional Intelligence): www.megaforte.com.sg/

SOUTH AFRICA
Department of Education Sciences, Rand Afrikaans University (Helen Krige):
http://general.rau.ac.za/eduscie/krige.htm

SWEDEN
Social Emotional Training (Social Emotionell Träning): www.set.st/

TURKEY
Emotional Intelligence of Turkey (Türkiyenin Duygusal Zekasi): www.duygusalzeka.com/

UNITED KINGDOM—ENGLAND
Ei United Kingdom: www.eiuk.com
The Mental Health Foundation
http://www.mentalhealth.org.uk/

UNITED KINGDOM—SCOTLAND
Promoting Social Competence Project: www.dundee.ac.uk/psychology/prosoc.htm

WEB LISTINGS FOR SOCIAL-EMOTIONAL LEARNING PROGRAMMES WITH INTERNATIONAL PRESENCE

- www.researchpress.com—I Can Problem Solve (ICPS);
- www.quest.edu—Skills for Adolescence; Skills for Action; Violence Prevention;
- www.channing-bete.com—Promoting Alternative Thinking Strategies (PATHS);
- www.esrnational.org—Resolving Conflict Creatively Program (RCCP);
- www.responsiveclassroom.org—Responsive Classroom;
- www.cfchildren.org—Second Step;
- www.peaceeducation.com—Peace Works;
- www.open-circle.org —Open Circle/Reach Out to Schools Social Competency Program;
- www.umdnj.edu/spsweb; www.EQParenting.com—Social Decision Making/Social Problem Solving Program;
- www.tribes.com—Tribes TLC: A New Way of Learning and Being Together.

RESOURCES FOR SERVICE LEARNING/CITIZENSHIP EDUCATION

- International Partnership for Service-Learning: www.ipsl.org
- National Center for Learning and Citizenship: www.ecs.org/clc
- Center for Information and Research on Civic Learning and Engagement: www.civicyouth.org
- National Service-Learning Exchange: www.nslexchange.org
- National Service-Learning Clearinghouse: www.service-learning.org
- International Education and Resource Network: www.iEARN.org

iEARN is a non-profit global network that enables young people to use the Internet and other new technologies to engage in collaborative educational projects that both enhance learning and make a difference in the world.

INTERNATIONAL RESOURCES FOR PARENTS

Listing of publishers of international editions of Elias, M.J.; Tobias, S.E.; Friedlander, B.S. 2000. *Emotionally intelligent parenting: How to raise a self-disciplined, responsible, socially skilled child.*
Australia: Transworld
Germany: Ullstein Buchverlage

Indonesian:	Mizan Publishers
Israel:	Matar-Triwaks Enterprises
Italy:	Newton & Compton Editorial
Netherlands:	Uitgeverij Het Spectrum
Polish:	Wydawnictwo Moderski
Portuguese:	Editora Objetiva (Brazil Only)
Portuguese:	Editora Pergaminho Lda (Portugal Only)
Rep. of Korea:	Segyesa Publishing Co.
Romanian:	Curtea Veche
Spain:	Random House Mondadori/Plaza and Janes
Thailand:	Plan Booknet/Plan for Kids
UK:	Hodder & Stoughton
United States:	Random House/Three Rivers Press

http://www.temple.edu/LSS/upcoming—Laboratory for Student Success/School-Family Partnership Project

PRIMARY WEB SITES FOR PUBLICATIONS, MATERIALS AND MEDIA

Collaborative for Academic, Social, and Emotional Learning: www.casel.org
George Lucas Educational Foundation: www.GLEF.org
Character Education Partnership: www.character.org
National Professional Resources: www.nprinc.com
Center for Social-Emotional Education: www.csee.net
The Harnessing Optimism and Potential Through Education Foundation: www.communitiesofhope.org.

Previous titles in the 'Educational practices series':

1. Teaching *by Jere Brophy.* 36 p.
2. Parents and learning *by Sam Redding.* 36 p.
3. Effective educational practices *by Herbert J. Walberg and Susan J. Paik.* 24 p.
4. Improving student achievement in mathematics *by Douglas A. Grouws and Kristin J. Cebulla.* 48 p.
5. Tutoring *by Keith Topping.* 36 p.
6. Teaching additional languages *by Elliot L. Judd, Lihua Tan and Herbert J. Walberg.* 24 p.
7. How children learn *by Stella Vosniadou.* 32 p.
8. Preventing behaviour problems: what works *by Sharon L. Foster, Patricia Brennan, Anthony Biglan, Linna Wang and Suad al-Ghaith.* 30 p.
9. Preventing HIV/AIDS in schools, *by Inon I. Schenker and Jenny M. Nyirenda,* 32 p.
10. Motivation to learn, *by Monique Boekaerts,* 28 p.

These titles can be downloaded from the websites of the IEA (http://www.curtin.edu.au/curtin/dept/smec/iae) or of the IBE (http://www.ibe.unesco.org/International/Publications/pubhome.htm) or paper copies can be requested from: IBE, Publications Unit, P.O. Box 199, 1211 Geneva 20, Switzerland. Please note that no. 2 and no. 4 are out of print, but can be downloaded from the websites.

The International Bureau of Education—IBE

An international centre for the content of education, the IBE was founded in Geneva in 1925 as a private institution. In 1929, it became the first intergovernmental organization in the field of education. In 1969, the IBE joined UNESCO as an integral, yet autonomous, institution.

It has three main lines of action: (a) organizing the sessions of the International Conference on Education; (b) collecting, analysing and disseminating educational documentation and information, in particular on innovations concerning curricula and teaching methods; and (c) undertaking surveys and studies in the field of comparative education. At the present time, the IBE: (a) manages *World data on education*, a databank presenting on a comparative basis the profiles of national education systems; (b) organizes regional courses on curriculum development; (c) collects and disseminates through its databank INNODATA notable innovations on education; (d) co-ordinates preparation of national reports on the development of education; (e) administers the Comenius Medal awarded to outstanding teachers and educational researchers; and (f) publishes a quarterly review of education—*Prospects*, a newsletter—*Educational innovation and information*, as well as other publications.

In the context of its training courses on curriculum development, the Bureau is establishing regional and sub-regional networks on the management of curriculum change and developing a new information service—a platform for the exchange of information on content.

The IBE is governed by a Council composed of representatives of twenty-eight Member States elected by the General Conference of UNESCO. The IBE is proud to be associated with the work of the International Academy of Education and publishes this material in its capacity as a clearinghouse promoting the exchange of information on educational practices.

http://www.ibe.unesco.org

Motivation to learn

By Monique Boekaerts

IBE

UNESCO

The International Academy of Education

The International Academy of Education (IAE) is a not-for-profit scientific association that promotes educational research, its dissemination, and the implementation of its implications. Founded in 1986, the Academy is dedicated to strengthening the contributions of research, solving critical educational problems throughout the world, and providing better communication among policy makers, researchers and practitioners. The seat of the Academy is at the Royal Academy of Science, Literature and Arts in Brussels, Belgium, and its co-ordinating centre is at Curtin University of Technology in Perth, Australia.

The general aim of the IAE is to foster scholarly excellence in all fields of education. Towards this end, the Academy provides timely syntheses of research-based evidence of international importance. The Academy also provides critiques of research, its evidentiary basis, and its application to policy.

The current members of the Board of Directors of the Academy are:

- Erik De Corte, University of Leuven, Belgium (*President*)

- Herbert Walberg, University of Illinois at Chicago, United States of America (*Vice President*)

- Barry Fraser, Curtin University of Technology, Australia (*Executive Director*)

- Jacques Hallak, Paris, France

- Michael Kirst, Stanford University, United States of America

- Ulrich Teichler, University of Kassel, Germany

- Margaret Wang, Temple University, United States of America

  http://www.curtin.edu.au/curtin/dept/smec/iae

Preface

This booklet explains principles that encourage children to learn and has been prepared for inclusion in the Educational Practices Series developed by the International Academy of Education and is distributed by the International Bureau of Education and the Academy. As part of its mission, the Academy provides timely syntheses of research on educational topics of international importance. This booklet is the tenth in the series on educational practices that improve learning. It opens a new door, however, since it focuses on behaviour rather than academic learning.

The author of this booklet, Monique Boekaerts, began her career as a teacher but decided to take up the study of psychology to understand better what went on in the minds of her students. She is a full professor at Leiden University in the Netherlands and has published over 120 papers and book chapters on motivation and self-regulation. She set up collaborative innovation programmes with the school-management and teachers of large vocational schools. Together with Teaching and School Management Consultants (TSM) she coaches the change processes that are currently taking place in vocational education. Professor Boekaerts has served as president of the European Association for Research in Learning and Instruction.

The officers of the International Academy of Education are aware that this booklet is based on research carried out primarily in economically advanced countries. The booklet, however, focuses on aspects of learning and behaviour that that may be found in most cultures in varying degrees. The principles presented here are likely to be generally applicable and useful throughout the world. Even so, the principles should be assessed with reference to local conditions, and adapted accordingly. In any educational setting, nation or culture, suggestions or guidelines for practice require sensitive and sensible application, and continuing evaluation.

HERBERT J. WALBERG
Editor, IAE Educational Practices Series
University of Illinois at Chicago

Table of contents

Introduction

In the last forty years, researchers have studied student motivation and have learned a great deal about:
- What moves students to learn and the quantity and quality of the effort they invest;
- What choices students make;
- What makes them persist in the face of hardship;
- How student motivation is affected by teacher practices and peer behaviour;
- How motivation develops;
- How the school environment affects it.

Most of the motivation research focused on well-adjusted students who are successful in school. However, successful students differ from their less-successful peers in many ways. For example, they often have clear ideas of what they want and do not want to achieve in life. Moreover, they perceive many learning settings as supportive of their own wishes, goals and needs, and react positively to the teacher's motivational practices.

This booklet is a synthesis of principles of motivation that have emerged from research into the effect of motivational practices on school learning. It addresses more traditional aspects, such as achievement motivation, intrinsic motivation and goal orientation, as well as the effect of teacher practices that promote motivational beliefs, motivation strategies and willpower. It focuses on learning goals and the effect of motivation on the pursuit of these goals, whilst recognizing the need for teacher practices that target socio-emotional goals as well.

Much of the research supporting the principles specified in this booklet stems from studies that investigated the association between motivation (seen as a student characteristic) and learning outcomes. Other principles have their origins in the theory of self that children and adolescents themselves develop through the years. Still other principles are based on research that showed how the opportunities that teachers and schools provide for learning and personal development (instructional procedures, teacher behaviour and classroom climate) are congruent or in conflict with the students' needs and goals. Priority was given to those principles that teachers can apply in their classrooms. It is the aim of this short introduction to motivation to make teachers aware that youngsters'

psychological needs change continuously. They change not just as a function of their developing knowledge and expertise in a particular subject-matter domain, but also in relation to their emerging theory of self in relation to that domain.

In this booklet, the reader will get to know two youngsters, namely Stefano and Sandra, who are both 11 years old and are attending school in different parts of the world. Stefano is the son of a car mechanic. He goes to school in a rural area in the south of Europe. Sandra is the daughter of a road worker. She attends school in a big city in South America. It is my intention to describe the thoughts, feelings and actions of these two children in order to provide an illustration of the various constructs described in the research sections. I hope that teachers will perceive these students' developing values, interests and goals as similar to what they actually observe in their own classrooms.

The eight principles addressed in this booklet are meant to be understood as pieces in a jig-saw puzzle that fit together to provide a coherent, comprehensive picture of how to provide a powerful environment for motivation strategies to develop. If you want to find out more about these eight principles, or about a specific principle, you can consult the literature on motivation. References are provided in relation to each principle.

1. Motivational beliefs

Motivational beliefs act as favourable contexts for learning.

Research findings

In the classroom the content covered and the social context vary continuously. Hence, children are frequently involved in unfamiliar learning situations. This may create ambiguity and uncertainty for some students and challenge for other students. Students try to make sense of novel learning situations by referring to their motivational beliefs. Motivational beliefs refer to the opinions, judgements and values that students hold about objects, events or subject-matter domains. Researchers have described the beliefs that students use to assign meaning to learning situations. A specific set of motivational beliefs pertains to the value students attach to a domain. For example, Stefano often says: 'I cannot see what I can possibly learn from reading poetry;' while Sandra states: 'Reading poems is the nicest activity we do at school.'

Motivational beliefs also refer to the student's opinion of the efficiency or effectiveness of learning and teaching methods (Stefano: 'Why do we always have to work in groups? I can learn better when I work alone'). Beliefs about internal control can be distinguished into self-efficacy beliefs and outcome expectations. Self-efficacy beliefs are opinions that students hold about their own ability in relation to a specific domain (Stefano: 'I believe that I am good at solving this type of mathematics problem;' Sandra: 'I am not a star in math, but I know how to analyze a reading text'). Outcome expectations are beliefs about the success or failure of specific actions (Stefano: 'I have been working at this grammar task for a long time and I still cannot get it right. I am certain I will not be able to come up with an acceptable solution').

Research has indicated that motivational beliefs result from direct learning experiences (e.g. Sandra: 'Most math problems are too difficult for me to get them right the first time. However, when somebody gives me a hint I can solve a lot of problems'),

observation learning (e.g. Stefano: 'The math teacher gets annoyed when students do not offer help to each other'), verbal statements by teachers, parents or peers (e.g. Sandra: 'My father thinks it is nonsense to learn poetry in school; he says mathematics is far more important') and social comparisons (e.g. Stefano: 'Why do I always get scolded, while the teacher never says anything to other students?').

Motivational beliefs act as a frame of reference that guides students' thinking, feelings and actions in a subject area. For example, motivational beliefs about mathematics determine which strategies students think are appropriate to do specific tasks. It is noteworthy that a student's beliefs about a domain may be dominantly favourable (optimistic) or unfavourable (pessimistic), thus providing a positive or negative context for learning. Once formed, favourable and unfavourable motivational beliefs are very resistant to change.

Motivating your students

As teachers, you should have a good idea of the motivational beliefs that your students bring into the classroom. It is important that you are aware that your students may already have formed favourable or unfavourable beliefs about a topic before they come into class. Knowledge about your students' motivational beliefs will allow you to plan learning activities that make good use of their favourable motivational beliefs and prompt them to reconsider unfavourable beliefs. Students are very successful in hiding their thoughts and feelings, leading to misconceptions about their values, self-efficacy beliefs and outcome expectations.

The set of principles addressed in this booklet will hopefully provide more insight into students' motivational beliefs and into the way these beliefs affect their involvement, commitment and engagement in the life classroom. Knowledge of these principles will, I hope, act as guidelines for helping students to establish favourable motivational beliefs and unmask unfavourable beliefs.

References: Pintrich, 2001; Skinner, 1995; Stipek, 1988; Vermeer, Boekaerts & Seegers, 2000.

2.   Unfavourable motivational beliefs impede learning

Students are not motivated to learn in the face of failure.

Research findings

Fear of failure does not automatically lead to passivity or avoidance. What matters are the motivational beliefs that have been attached to a subject-matter area. For example, Stefano has dominantly favourable beliefs about mathematics and unfavourable beliefs in relation to language learning. Domain-specificity of motivational beliefs implies that a student may be failure-oriented in some domains and not in others. Stefano no longer perceives a relationship between what he can do (his actions) and the outcomes of his actions (success or failure) in the language domain. He feels uncertain, stating that he is unable to perform the tasks well. Students give different reasons for their success or failure in various school subjects and these reasons are consistent with their self-concept of ability in that domain. The main reasons Stefano gives for his poor performance in languages is his lack of ability. Other frequently used excuses for poor performance are lack of effort (Sandra: 'I did poorly in history today because I did not put in a lot of effort'), bad luck (Stefano: 'I was unlucky that I was called upon first to consider that question'), inadequate strategy use (Stefano: 'I solved the math problem correctly, but I did not know that we had to write down the solution steps as well') and task characteristics (Sandra: 'The math problem was just too difficult'). Children who view poor performance as the result of low ability expect failure to occur again and again. These students experience negative thoughts and feelings (e.g. Sandra: 'I am the only one with seven mistakes. The teacher will not like me because I am a dumb kid'). Negative thoughts that are repeatedly associated with a task or activity become attached to similar learning situations. As such, a whole domain may be categorized as 'too difficult' or 'threatening'. Once these unfavourable motivational beliefs have become part of a

student's theory of self, they will be activated again and again, creating doubt and anxiety. Unfavourable beliefs impede the learning process because they direct the learners' attention away from the learning activity itself, focusing it instead on their low ability. Even though children's understanding of causality changes with age, their beliefs about the cause of their successes and failures in a particular domain are very resistant to change.

Motivating your students

Students who state that they will never be able to complete the task successfully signal to you that they no longer perceive a link between their actions and a positive outcome. You can help them to re-establish the link by creating learning situations where they can experience success. However, it is not sufficient that they get the correct solution. They also need to understand why the solution plan was correct and what they can do (actions) to improve their skill further. Your students' attention has to be drawn explicitly to the link between their actions and the outcome of their actions by asking questions such as: 'What did you do to get that solution? How do you know that the strategy you used is effective? Would this strategy work for the following problem as well? Why or why not?'

Paradoxically, students who have established unfavourable motivational beliefs are not interested in such process-oriented feedback. They only want to know whether their answer is correct, or whether they are on the right track. Try to be alert when your students request outcome-related feedback. Focus on what they have already mastered (e.g. 'Stefano, you got three correct. That is better than yesterday.') rather than on their shortcomings. Better still, point out the strengths of their solution plan. Such process-oriented feedback gives them a feeling of progress, which is necessary to build up a positive identity as a successful learner. Gradually stimulate them to reflect on their own performance (self-assessment). For example, encourage Stefano to verbalize why the corrected sentence conveys his message better.

References: Covington, 1992; Stipek, 1988; Turner & Meyer, 1998; Vermeer et al., 2000; Ryan, Gheen & Midgley, 1998.

3. Favourable motivational beliefs facilitate learning

> Students who value the learning activity are less dependent on encouragement, incentives and reward.

Research findings

Students are more interested in doing activities for which they think they have the necessary competence, or that they value (e.g. Stefano: 'I like math because it is easy, and I need it to become a space engineer', or Sandra: 'I don't like math, but I do my best because my dad tells me that it is important'). Students who value new skills have established favourable motivational beliefs. The chances are good that they are interested in opportunities to practice these skills. It is important to distinguish such commitment from mere compliance with the teacher-set goals. Many students complete tasks that they do not value all that much simply because they expect some sort of reward (e.g. high marks, a pass, or social approval). Students who undertake learning tasks purely for the sake of getting a reward from others, or in order to avoid some penalty, are extrinsically motivated (e.g. Stefano: 'I hate grammar exercises, but my mother prepares my favourite meal when I have to study for a test'). An activity is generally considered to be intrinsically motivating if external reward is not necessary for students to initiate and continue that activity. Favourable motivational beliefs are attached to the activity itself. Students who are intrinsically motivated will report that they do not have to invest effort and that doing the activity is gratifying (e.g. Sandra: 'when I am writing poetry or stories for the school bulletin, I lose track of time'). When difficulties arise, these students will persist with the activity because they experience a feeling of self-determination.

Motivating your students

Unfortunately, not all students are intrinsically motivated and you also have to cater to those students who are less motivated to learn. It is important to realize that classroom climate and the way you interact with your students facilitates or impedes

their motivation. Try to make tasks and activities meaningful for your students by referring to the intrinsic value of the task and to potential applications in other subject areas and outside school. How can you help your students to develop favourable motivational beliefs? Translate the curriculum in terms of the skills that your students find relevant and interesting. Find out what their current interests and future career goals are (e.g., Sandra wants to become a nurse and Stefano wants to become a space engineer). Show a video, a newspaper cutting, or tell a story, highlighting the importance and functional relevance of new content and skills. Ask students who are already motivated to explain why they value these new skills. Alternatively, ask your students to interview their parents, other teachers in school or older students to find out when they use the new content or skills. These activities will catch your students' attention and curiosity. This is already half of the motivation story. The other half is holding their interest. It is important that students perceive an optimal match between perceived demands and their current capacity. Allow them to adapt exercises according to their current capacity. For example, Stefano gets bored when math problems are too easy. Do not force him to cover the content of the lesson at the same pace, or in the same way, as the slower learners. Also, encourage students who find a math problem too demanding, to redesign it in such a way that it becomes less threatening (e.g. Sandra: 'Can I do this math problem together with Claudia?'). Allowing students to adapt a learning activity to their own psychological needs gives them a feeling of autonomy and self-determination. Denying them this right will be interpreted as external pressure to comply.

References: Bruning & Horn, 2000; Guthrie & Solomon, 1997; Ryan & Deci, 2000; Stipek, 1988; Turner & Meyer, 1998; Wlodkowski & Jaynes, 1990.

4.   Students' beliefs about goal orientation

> Students who are mastery-oriented learn more than students who are ego-oriented.

Research findings

An important motivational belief that has not been discussed so far is goal orientation. The way students' orient themselves to learning tasks within a domain is a strong indicator of their engagement and performance. Students who learn because they want to master a new skill use more effective learning strategies than students who are ego-oriented. The latter students engage in learning tasks with the intention to demonstrate success (approach ego-orientation) or to hide failure (avoidance ego-orientation). The motivation process of mastery-oriented students differs from that of ego-oriented students in many ways. For example, Stefano shows mastery-orientation in relation to the math domain and ego-orientation in relation to language domain. He starts on his math home-work before dinner because he wants to find out whether he can solve the problems. He is prepared to invest effort because he values mathematics and enjoys improving his math skills. When Stefano meets obstacles while doing math, he asks himself: 'How can I make it work?' He is not ashamed that others hear about his mistakes. On the contrary, he always volunteers to show his solution plan, because he appreciates the feedback he gets. In contrast, Stefano does not want others to find out that he made many spelling and grammatical mistakes in a text.

Sandra also values mathematics but for different reasons. She is ego-oriented in math class. She wants to demonstrate success to change other people's opinion about her math ability. Sandra invests effort in math as long as she feels confident that she can find the correct solution. She gives up when she spots mistakes, because she believes that there is only one correct solution. These beliefs fuel her fear that others will use her mistakes as proof of her math ability.

Two research findings should be reported here. Firstly, students display a dominant goal orientation (ego or mastery) by the time they are in second grade, and striving for ego-orientation goals becomes more dominant as children proceed through primary school. They become progressively more concerned with their self-worth, express more concern for peer-status and avoid doing things that the group rejects (fear of alienation). By the fourth grade, avoidance ego goals (e.g. wanting to hide mistakes) have already assumed a prominent position. A second finding shows that teachers set up dominantly competitive or co-operative learning settings in class. Teachers who highlight evaluation procedures, give public feedback, frequently make social comparisons and refer to individual abilities create a competitive atmosphere and elicit ego-oriented thoughts and feelings.

Motivating your students

The extent to which you succeed in creating a mastery-oriented learning setting is an indication of your professional competence. You can play down ego-orientation by explaining to your students that you are not interested in seeing one correct outcome, but that you focus instead on their attempts to come up with a solution strategy. Students will only believe this 'trying is more important than the product' statement when you act according to what you preach. In other words, provide feedback with respect to the solution plan, encourage students to exchange information about the strategies they used and allow them to learn from their mistakes. This is a difficult job since ego-oriented students get annoyed when they have to reflect on their mistakes. By using supportive comments that highlight their involvement, progress and effort you will convince them that you value their attempts to solve problems, particularly when they reflect about what did not work out and why. Mastery-orientation will develop when these students take pride in finding parts of a solution and in catching errors in progress.

References: Elliot, 1999; Niemivirta, 1999; Pintrich, 2001; Turner & Meyer, 1998; Vermeer et al., 2000.

5. Different beliefs about effort affect learning intentions

Students expect value for effort.

Research findings

Students decide how much effort they will allocate to a learning task on the basis of their self-concept of ability and their effort beliefs. Young children are notorious over-estimators or under-estimators of their own performance. They may rate themselves among the best of their class, even though their performance is absolutely below the mark. Young children have a rather naïve theory of effort. They believe that if they want something badly enough and do their best to accomplish it, they will be valued for their effort. In other words, they think they have control over the learning situation and keep their high expectations of success even after repeated failure. Their conceptualization of effort as the most important explanation of their successes and failures is a strong motivator to keep practicing.

However, as students get older, the messages they receive from parents and teachers change gradually. More emphasis is put on their ability as a major source of success and failure than on their effort. Children learn to take into account their actual experiences and evaluative feedback from others. They also engage in social comparisons with their peers. This implies that their domain-specific self-efficacy beliefs become more accurate and realistic. Simultaneously, they link these beliefs to their emerging theory of effort. By the age of 9, children seem to have lost confidence in effort as the overall source of success. Research evidence is clear: domain-specific self-efficacy beliefs influence effort investment, and not the other way round. Students like Stefano, who believe that they are good in mathematics, are willing to invest effort to acquire math skills, but they do not necessarily invest more observable effort. Their task-engagement is fundamentally different from that of students

who believe they lack efficiency. More specifically, these students use adequate cognitive strategies that lead to good results. Students like Sandra, who believe that their math skills are deficient, may also invest effort in mathematics. However, they do a lot of things that are ineffective, such as sitting and sighing in front of their books, copying a lot of exercises, rereading several pages. This type of effort creates anxiety and frustration and leads to poor performance. Research has shown that teachers can coach students to develop their effort beliefs. Interestingly, teachers who coach effort are rewarded by enhanced intrinsic motivation.

Motivating your students

Teacher observations confirm that students develop a threshold for declaring whether or not they have put in sufficient effort to reach the learning goal. They use specific stop rules. For example, Sandra may say: 'I have worked for more than an hour now. This must be sufficient for my math homework', or 'I have worked harder for mathematics than for history'. Stefano may justify thus: 'I don't have to work hard for math, I just do the exercises and it usually works out well', or 'I have worked longer than any of my friends to write a good text—this must be sufficient'.

In general, students' theory of effort is underdeveloped. They need assignments to build up domain-specific effort beliefs and to be encouraged to update these beliefs as their skill develops. When you encourage and value effort, your students will begin to view themselves as responsible for their own learning. It is essential, however, that you provide your students with adequate feedback. A good way to start is by providing assignments that require students to predict the effort needed to do a task. After finishing the task, students could be asked to reflect on the invested effort. Was it sufficient or superfluous, and why? Once students get into the habit of reflecting on their effort, they are better equipped to self-regulate their own learning.

References: Boekaerts, 1997; Covington, 1992; Pintrich, 2001; Wlodkowski & Jaynes, 1990; Ryan & Deci, 2000.

6.   Goal setting and appraisal

Students need encouragement and feedback on how to develop motivational strategies.

Research findings

Students who define teacher-set goals in terms of their own reasons for learning create a commitment to a desired end-state. Their goal-setting process differs fundamentally from that of students who merely comply with the teacher's expectations. Recent findings indicate that learning goals that are agreed upon jointly by the students and the teacher have a better chance of being accomplished. Such an agreement reflects the intention of both parties to invest effort.

Setting a learning goal refers to the selection of a motivation strategy that fits the actual learning situation. This strategy consists of active attempts on the part of the learner to activate favourable motivational beliefs, to pay attention to relevant cues in the learning environment, and to ignore cues that are distracting from learning. Students who take the time to appraise learning situations in terms of their own goals discover desirable and undesirable end-states. For example, Stefano hated all exercises in which he had to use a dictionary. However, recognition of desirable outcomes of a language activity was a turning point in his attitude. His teacher recommended that he send a letter to a Scottish boy who wants to become a space engineer. Stefano's favourable appraisal of the pen-pal context and the anticipated desirable outcomes (getting an answer) turned him from a passive language learner into an active one. He learned to pay attention to positive outcomes and ignore undesired end-states (spelling mistakes), and he discovered the power of writing as a tool for communication.

Students who begin the learning process by activating favourable beliefs, particularly mastery-orientation and self-efficacy beliefs, need less encouragement from others to get started. Moreover, favourable motivational beliefs draw students' attention to cues in the environment that elicit further interest and confidence in their own capacity to do the task.

Motivating your students

Within the context of the classroom, the teachers' main goal is to get through the syllabus. Most teachers still overrate their students' capacity to set their own learning goals. Hardly any time or effort is devoted to obtaining the students' opinions about the relevance and value of the learning tasks. Consequently, students can motivate themselves for out-of-class activities but do not have a clue about how they can motivate themselves for their schoolwork. Yet, in the goal-setting phase, students lay the foundation for further learning and for the development of interest. What can be done to encourage your students to develop motivation strategies? The goal-setting process can be facilitated by asking students to stop and think about why a particular learning task is important, relevant, fun, boring, challenging, difficult or easy. Why are they confident (or doubtful) about their own skills to do a task, and what triggers their doubt or confidence? When students have completed a task they can reflect on their original appraisal of the task again. Ask them to formulate in their own words whether their appraisal of the task has changed and why. By asking your students to reflect on their initial competence and relevance judgements in relation to different learning tasks and about their initial outcome expectations, you create a favourable classroom climate for goal setting. Your students will feel free to make their appraisals explicit and open for discussion, raise questions about their own and other students' motivation for learning, and learn from each other. If you show interest in the reasons why your students consider some topics as their favourites while others find these topics boring, both you and your students will gain information about what makes motivation strategies work.

References: Boekaerts, 1997; Boekaerts, 2001; Niemivirta, 1999; Turner & Meyer, 1998; Vermeer, et al., 2000.

7. Striving for goals and willpower

> Students need encouragement and feedback on how to develop willpower.

Research findings

Good intentions that were strong in the goal-setting stage do not automatically lead to goal accomplishment. Many learning goals need active striving on the part of the learner in order to be accomplished, meaning that effort needs to be invested. Effort refers to an intentional act that increases commitment to a task, such as increasing attention, concentration and the amount of time spent on a task, or by doing specific activities (e.g. re-reading, rehearsal, underlining, paraphrasing, copying). However, effort often declines when a task gets more complex or less interesting, when obstacles are encountered, or when students are distracted by competing activities. At such a point, they need willpower to sustain attention and effort.

Parents and teachers alike view persistence as an important aspect of willpower. Yet, research has shown that persistence is not necessarily a virtue. Some students try the same strategy again and again in order to complete a task (high persistence) while others discard a strategy at the first sign of failure (low persistence). Results from recent studies suggest that two important learning strategies should be implemented. The first strategy deals with the students' capacity to initiate a solution plan without too much hesitation. The second strategy deals with the students' capacity to judge whether it is fruitful to continue with a solution plan (persistence), or whether it is better to give it up because it will lead nowhere (disengagement).

Before initiating a learning activity, students should orient themselves to the learning task in terms of its purpose and possible solution plans. Effective decisions to persist in the goal-striving stage are based on this knowledge. Students who have a good conception of the learning goal and also have access to a repertoire of strategies to generate an adequate solution plan

use their effort constructively. They can judge which strategies are useful and also monitor whether the selected strategies are effective to reach the goal. If they notice that a chosen strategy is not effective, they can select a new one and test whether it is more effective or else disengage from the task because they judge that effort is no longer fruitful (e.g. not enough time or resources). Students who have a misconception of the goal or lack adequate strategies may also persist, but their effort is largely undirected. For example, Sandra often tries several solution plans blindly when she is doing her math homework in the hope that one will work.

Motivating your students

How can you help your students to develop willpower? First of all, you should not be misled by observed effort. When effort investment is high (or low), you still need to know why that is the case. In order to be able to interpret student initiative, persistence and disengagement meaningfully, you need to have a good idea of the way your students perceive the learning goal and also of how much effort they need to invest to reach it. Students should be given plenty of opportunities to practice striving for goals. You can coach this process by reminding them to set a series of sub-goals and to compose a checklist that will help them to monitor, assess and reflect on the quality of their engagement and commitment during the solution process.

Reflecting on the goal-striving process implies that students should raise questions about the resources that are necessary and sufficient to reach various sub-goals. For example, Stefano may ask himself: 'Do I have sufficient time to finish my history homework before dinner if I reread every section twice and make a brief summary?' Post-activity reflection about effort investment is essential to make students aware of their attempts at effort management and of the reason why they did not exercise willpower. By asking your students to compare and contrast the amount and type of effort invested in various tasks, you can help them to develop their theory of effort, and at the same time allow them to gain insight into their own willpower.

References: Boekaerts, 1997; Boekaerts, 2001; Corno & Randi, 1997; Niemivirta, 1999; Skinner, 1995.

8.   Keeping multiple goals in harmony

> Students are more committed to learning if the objectives are compatible with their own goals.

Research findings

Teachers, educators and parents are convinced that acquiring new knowledge and skills is the most important goal that students should strive for in a school context. The reality is different. Youngsters do not consider the learning goals set by the teacher as the most salient goals in their life. They pursue many other goals as well. For example, they want to be treated fairly, build up a network of friends, learn more about their favourite topics and discuss romantic partners. These personal goals play a crucial role in motivation processes by defining their content, direction and intensity. Recent evidence suggests that students are more motivated towards their schoolwork when school-related goals are in harmony with their own wishes, needs and expectations. For instance, Sandra adores her teacher and uses her as a role model because she acknowledges that Sandra wants to become a nurse and frequently relates schoolwork to this important goal. Students who note that the teacher acknowledges their personal goals accept the teacher's goals more easily. By contrast, students who realize that their personal goals are ignored, or even thwarted, rebel against the system and consider the curriculum as alien to their 'real' life.

Teachers and parents often complain that students do not adopt the goals they hold for them, and that they do not follow up on their well-meant advice. For example, Stefano's father tries to prevent him from doing his homework with the radio on, believing that music affects motivation and performance negatively. Current research does not support this view. Yet, such conflicts of interest lead to the frustration of Stefano's need for autonomy. Often, teachers (and parents) try to push their own goals along, thus fueling the child's struggle for autonomy. For decades, schools, teachers and researchers narrowed educa-

tional goals to learning and achievement, which only frustrated students' social goals.

Motivating your students

Students bring their own goals into the classroom and want to negotiate with you about how, when, and with whom they want to reach the learning goals. It is important to realize that you impose many goals on your students, including social goals (e.g. 'You have to work individually, without the support or help from your peers'; or 'You have to work in small groups and take responsibility for the learning of members of your group'). Peers also impose goals on other students (e.g. 'Ignore the teacher when he asks for volunteers'). When students realize that their own goals are discordant with your goals, they make attempts to align the curricular goals with their own goals. For example, Sandra may ask: 'Can I hand in my homework tomorrow because I did not have enough resource material to make a good job of it?' Similarly, Stefano may request: 'Can I do this task alone, because I have a different opinion than the rest of my group?' If you grant these requests, your students will experience self-determination. The positive cognitions and feelings that are part of that experience will further the learning process. On the contrary, if you deny these requests, they will experience a conflict of goals and may not take responsibility for achieving the curricular goals. Many forms of misbehaviour in class can be interpreted in terms of a goal conflict. You will deal more flexibly with misbehaviour when you view it as a signal that a salient goal is being frustrated. For example, Stefano may say: 'How can I work efficiently on a math problem if you want me to help students who always run into problems?' Likewise, Sandra may ask: 'Why can't we do this task together?' It is important to realize that your students want to be treated with respect. They expect you to explain why you turn down their requests.

References: Boekaerts, 1998; Boekaerts, 1999; Maehr, 1984; Wentzel, 1996.

Conclusion

It is often stated that bad teaching kills motivation and that good teaching brings out the best in students of all ages. If you want to encourage your students to become their own teachers and develop independent learning skills, you need to know about the principles that guide motivated learning. The eight principles that are addressed in this booklet apply to children and adolescents from different countries and different cultures. I described the principles in such a way that you gain insight into the reasons why students are or are not motivated to learn in the context of the classroom. However, you still need to adapt these principles to the local context of your classroom. I focused on two primary school students, Stefano and Sandra, and referred to their thinking and feeling in relation to the mathematics and language domains, yet the principles do not refer to particular curricula or specific age groups. Rather, they refer to generic aspects of motivated learning that cut across school subjects, grade levels and types of education. They focus on the students' beliefs, opinions and values and how these motivational beliefs affect learning. Knowledge of your students' motivational beliefs will help you to create learning environments that are well suited to their psychological needs. The capacity to listen to your students and observe their behaviour in the live classroom will help to inform you of what they find interesting, challenging, boring and threatening, and why they have this opinion. Willingness to negotiate with your students and grant them autonomy will convince them that you are truly interested in how and why they learn. A good way to start your observations is by selecting one or more students in your class who think, feel and behave somewhat like Stefano or Sandra. Observe these students in the next few weeks and discover how the eight motivational principles that are described in this booklet work in *your* classroom.

References

Boekaerts, M. 1997. Self-regulated learning: a new concept embraced by researchers, policy makers, educators, teachers, and students. *Learning and instruction* (Tarrytown, NY), vol. 7, no. 2, p. 151–86.

——. 1998. Boosting students' capacity to promote their own learning: a goal theory perspective. *Research dialogue in learning and instruction* (Exeter, UK), vol. 1, no. 1, p. 13–22.

——. 1999. Coping in context: goal frustration and goal ambivalence in relation to academic and interpersonal goals. *In*: Frydenberg, E., ed. *Learning to cope: developing as a person in complex societies,* p. 175–97. Oxford, UK, Oxford University Press.

——. 2001. Pro-active coping: meeting challenges and achieving goals. *In*: Frydenberg, E., ed. *Beyond coping: meeting goals, visions and challenges.* Oxford, UK, Oxford University Press.

Bruning, R.; Horn, C. 2000. Developing motivation to write. *Educational psychologist* (Hillsdale, NJ), vol. 35, no. 1, p. 25–37.

Corno, L.; Randi, J. 1997. Motivation, volition and collaborative innovation in classroom literacy. *In*: Guthrie, J.; Wigfield, A., eds. *Reading, engagement: motivating readers through integrated instruction*, p. 14–31. Newark, DE, International Reading Association.

Covington, M.V. 1992. *Making the grade: a self-worth perspective on motivation and school reform.* Cambridge, UK; New York, Cambridge University Press.

Elliot, A.J. 1999. Approach and avoidance motivation and achievement goals. *Educational psychologist* (Mahwah, NJ), vol. 34, no. 3, p. 169–89.

Guthrie, J.T.; Solomon, A. 1997. Designing contexts to increase motivations for reading. *Educational psychologist* (Mahwah, NJ), vol. 32, no. 2, p. 95–103.

Maehr, M.L. 1984. Meaning and motivation: toward a theory of personal investment. *In*: Ames, R.E.; Ames, C., eds. *Research on motivation in education: Vol. 1. Student motivation*, p. 115–44. San Diego, CA, Academic Press.

Niemivirta, M. 1999. Motivational and cognitive predictors of goal setting and task performance. *International*

journal of educational research (Oxford, UK), vol. 31,
p. 499–513.

Pintrich, P.R. 2001. The role of goal orientation in self-regula-
ted learning. *In*: Boekaerts, M.; Pintrich, P.R.; Zeidner,
M., eds. *Handbook of self-regulation*, p. 451–502. San
Diego, CA, Academic Press.

Ryan, R.M.; Deci, E.L. 2000. Self-determination theory and the
facilitation of intrinsic motivation, social development,
and well-being. *American psychologist* (Washington,
DC), vol. 55, p. 68–78.

Ryan, A.M.; Gheen, M.H.; Midgley, C. 1998. Why some students
avoid asking for help: an examination of the interplay
among students' academic efficacy, teachers' social-
emotional role, and the classroom goal structure. *Journal
of educational psychology* (Washington, DC), vol. 90,
no. 3, p. 528–35.

Skinner, E.A. 1995. *Perceived control, motivation and coping*.
Thousand Oaks, CA, Sage Publications.

Stipek, D.J. 1988. *Motivation to learn: from theory to practice*.
Englewood Cliffs, NJ, Prentice Hall.

Turner, J.C.; Meyer, D.K. 1998. Integrating classroom context
into motivation theory and research: rationales, methods,
and implications. *In*: Urdan, T.; Maehr, M.; Pintrich, P.,
eds. *Advances in motivation and achievement: a research
annual, vol. 11*, p. 87–121. Greenwich, CT, JAI Press.

Vermeer, H.; Boekaerts, M.; Seegers, G. 2000. Motivational and
gender differences: sixth-grade students' mathematical
problem-solving behavior. *Journal of educational psycho-
logy* (Washington, DC), vol. 92, no. 2, p. 308–15.

Wentzel, K.R. 1996. Social and academic motivation in middle
school: concurrent and long-term relations to academic
effort. *Journal of early adolescence* (Thousand Oaks,
CA), vol. 16, no. 4, p. 390–406.

Wlodkowski, R.J.; Jaynes, J.H. 1990. *Eager to learn*. San
Francisco, CA, Jossey Bass Publishers.

NOTES

Teaching

By Jere Brophy

The International Bureau of Education–IBE

An international centre for the content of education, the IBE was founded in Geneva in 1925 as a private institution. In 1929, it became the first intergovernmental organization in the field of education. In 1969, the IBE joined UNESCO as an integral, yet autonomous, institution with three main lines of action: organizing the sessions of the International Conference on Education; collecting, analysing and disseminating educational documentation and information, in particular on innovations concerning curricula and teaching methods; and undertaking surveys and studies in the field of comparative education.

At the present time, the IBE: (a) manages *World data on education*, a databank presenting on a comparative basis the profiles of national education systems; (b) organizes courses on curriculum development in developing countries; (c) collects and disseminates through its databank INNODATA notable innovations on education; (d) co-ordinates preparation of national reports on the development of education; (e) administers the Comenius Medal awarded to outstanding teachers and educational researchers; and (f) publishes a quarterly review of education–*Prospects*, a quarterly newsletter–*Educational innovation and information*, a guide for foreign students–*Study abroad*, as well as other publications.

In the context of its training courses on curriculum development, the Bureau is establishing regional and subregional networks on the management of curriculum change and developing a new information service–a platform for the exchange of information on content.

The IBE is governed by a Council composed of representatives of twenty-eight Member States elected by the General Conference of UNESCO.

The IBE is proud to be associated with the work of the International Academy of Education and publishes this material in its capacity as a clearing house promoting the exchange of information on educational practices.

http://www.ibe.unesco.org

Series preface

This booklet on the generic aspects of effective teaching has been prepared for inclusion in the Educational Practices Series developed by the International Academy of Education and distributed by the International Bureau of Education and the Academy. One mission of the International Academy of Education is to foster scholarly excellence in all fields of education. As part of this mission, the Academy provides timely syntheses of research on educational topics of international importance. This booklet is the first in a series on educational practices that generally improve learning. It focuses on the most central act of education–teaching.

The author is Jere Brophy, who is University Distinguished Professor of Teacher Education at Michigan State University and a Fellow of the International Academy of Education. He is well known both for his personal contributions to educational research and for his policy-oriented syntheses of work on various aspects of classroom teaching. He was one of the developers of process/product research, which examines relationships between teaching practices and student outcomes. Also, he has contributed to research and scholarship concerning teachers' attitudes, beliefs and expectations, including self-fulfilling prophecy effects; the interpersonal dynamics of teacher/student interaction; classroom management; student motivation; the analysis of instructional materials and learning activities; and the teaching of school subjects for understanding, appreciation and life application.

The Academy is grateful to Professor Brophy for planning, drafting and revising this booklet. Professor Brophy wishes to thank Lorin Anderson, Erik De Corte, Barry Fraser and Herbert Walberg for their comments on previous drafts of the booklet, and June Benson for her assistance with manuscript preparation.

The officers of the International Academy of Education are aware that this booklet is based on research carried out primarily in economically advanced countries. The booklet, however, focuses on aspects of teaching that appear to be universal in much formal schooling and thus seem likely to be generally applicable throughout the world. Even so, the principles need to be assessed with reference to local conditions,

and adapted accordingly. In any educational setting, guidelines for practice require sensitive and sensible application and continuing evaluation of their effectiveness.

<div align="right">

HERBERT J. WALBERG
Editor, Educational Practices Series
University of Illinois at Chicago

</div>

Officers of the International Academy of Education

- Erik De Corte, President of the Academy and Professor of Education, Catholic University of Leuven, Belgium.

- Herbert J. Walberg, Vice-President of the Academy and Research Professor of Education and Psychology, University of Illinois at Chicago, United States.

- Barry J. Fraser, Executive Officer of the Academy and Professor of Education, Curtin University of Technology, Perth, Australia.

Table of contents

Introduction

This booklet is a synthesis of principles of effective teaching that have emerged from research in classrooms. It addresses generic aspects of curriculum, instruction and assessment, as well as classroom organization and management practices that support effective instruction. It focuses on learning outcomes but with recognition of the need for a supportive classroom climate and positive student attitudes towards schooling, teachers and classmates.

Much of the research support for these principles comes from studies of relationships between classroom processes (measured through observation systems) and student outcomes (most notably, gains in standardized achievement tests). However, some principles are rooted in the logic of instructional design (e.g. the need for alignment among a curriculum's goals, content, instructional methods and assessment measures). In addition, attention was paid to emergent theories of teaching and learning (e.g. socio-cultural, social constructivist) and to the standards statements circulated by organizations representing the major school subjects. Priority was given to principles that have been shown to be applicable under ordinary classroom conditions and associated with progress towards desired student outcomes.

The principles rest on a few fundamental assumptions about optimizing curriculum and instruction. First, school curricula subsume different types of learning that call for different types of teaching, and so no single teaching method (e.g. direct instruction, social construction of meaning) can be the method of choice for all occasions. An optimal programme will feature a mixture of instructional methods and learning activities.

Second, within any school subject or learning domain, students' instructional needs change as their expertise develops. Consequently, what constitutes an optimal mixture of instructional methods and learning activities will evolve as school years, instructional units and even individual lessons progress.

Third, students should learn at high levels of mastery yet progress through the curriculum steadily. This implies that, at any given time, curriculum content and learning activities need to be difficult enough to challenge students and extend their

learning, but not so difficult as to leave many students confused or frustrated. Instruction should focus on the zone of proximal development, which is the range of knowledge and skills that students are not yet ready to acquire on their own but can acquire with help from their teachers.

1.   A supportive classroom climate

> **Students learn best within cohesive and caring learning communities.**

Research findings

Productive contexts for learning feature an ethic of caring that pervades teacher/student and student/student interactions and transcends gender, race, ethnicity, culture, socio-economic status, handicapping conditions and all other individual differences. Students are expected to manage instructional materials responsibly, participate thoughtfully in learning activities, and support the personal, social and academic well-being of all members of the classroom community.

In the classroom

To create a climate for moulding their students into a cohesive and supportive learning community, teachers need to display personal attributes that will make them effective as models and socializers: a cheerful disposition, friendliness, emotional maturity, sincerity, and caring about students as individuals as well as learners. The teacher displays concern and affection for students, is attentive to their needs and emotions, and socializes them to display these same characteristics in their interactions with one another.

In creating classroom displays and in developing content during lessons, the teacher connects with and builds on the students' prior knowledge and experiences, including their home cultures. Extending the learning community from the school to the home, the teacher establishes and maintains collaborative relationships with parents and encourages their active involvement in their children's learning.

The teacher promotes a learning orientation by introducing activities with emphasis on what students will learn from them, treating mistakes as natural parts of the learning process, and

encouraging students to work collaboratively and help one another. Students are taught to ask questions without embarrassment, to contribute to lessons without fear of their ideas being ridiculed, and to collaborate in pairs or small groups on many of their learning activities.

References: Good & Brophy (2000); Sergiovanni (1994).

2. Opportunity to learn

> Students learn more when most of the available time is allocated to curriculum-related activities and the classroom management system emphasizes maintaining their engagement in those activities.

Research findings

A major determinant of learning in any academic domain is the degree of exposure to the domain at school. The lengths of the school day and the school year create upper limits on students' opportunities to learn. Within these limits, the learning opportunities actually experienced by students depend on how much of the available time they spend participating in lessons and learning activities. Effective teachers allocate most of the available time to activities designed to accomplish instructional goals.

Research indicates that teachers who approach management as a process of establishing an effective learning environment tend to be more successful than teachers who emphasize their roles as disciplinarians. Effective teachers do not need to spend much time responding to behaviour problems because they use management techniques that elicit students' co-operation and sustain their engagement in activities. Working within the positive classroom climate implied by the principle of a learning community, the teacher articulates clear expectations concerning classroom behaviour in general and participation in lessons and learning activities in particular, teaches procedures that foster productive engagement during activities and smooth transitions between them, and follows through with any needed cues or reminders.

In the classroom

There are more things worth learning than there is time available to teach them, and so it is essential that limited classroom time be used efficiently. Effective teachers allocate most of this

time to lessons and learning activities rather than to non-academic pastimes that serve little or no curricular purpose. Their students spend many more hours each year on curriculum-related activities than do students of teachers who are less focused on instructional goals.

Effective teachers convey a sense of the purposefulness of schooling and the importance of getting the most out of the available time. They begin and end lessons on time, keep transitions short, and teach their students how to get started quickly and maintain focus when working on assignments. Good planning and preparation enable them to proceed through lessons smoothly without having to stop to consult a manual or locate an item needed for display or demonstration. Their activities and assignments feature stimulating variety and optimal challenge, which help students to sustain their task engagement and minimize disruptions due to boredom or distraction.

Successful teachers are clear and consistent in articulating their expectations. At the beginning of the year they model or provide direct instruction in desired procedures if necessary, and subsequently they cue or remind their students when these procedures are needed. They monitor the classroom continually, which enables them to respond to emerging problems before they become disruptive. When possible, they intervene in ways that do not disrupt lesson momentum or distract students who are working on assignments. They teach students strategies and procedures for carrying out recurring activities such as participating in whole-class lessons, engaging in productive discourse with classmates, making smooth transitions between activities, collaborating in pairs or small groups, storing and handling equipment and personal belongings, managing learning and completing assignments on time, and knowing when and how to get help. The teachers' emphasis is not on imposing situational control but on building students' capacity for managing their own learning, so that expectations are adjusted and cues, reminders and other managerial moves are faded out as the school year progresses.

These teachers do not merely maximize 'time on task', but spend a great deal of time actively instructing by elaborating content for students and helping them to interpret and respond to it. Their classrooms feature more time spent in interactive discourse and less time spent in solitary seatwork. Most of their instruction occurs during interactive discourse with students rather than during extended lecture presentations.

Note: The principle of maximizing opportunity to learn is not meant to imply maximizing the scope of the curriculum (i.e. emphasizing broad coverage at the expense of depth of development of powerful ideas). The breadth/depth dilemma must be addressed in curriculum planning. The point of the opportunity-to-learn principle is that, however the breadth/depth dilemma is addressed and whatever the resultant curriculum may be, students will make the most progress towards intended outcomes if most of the available classroom time is allocated to curriculum-related activities.

Note: Opportunity to learn is sometimes defined as the degree of overlap between what is taught and what is tested. This definition can be useful if both the curriculum content and the test content reflect the major goals of the instructional programme. Where this is not the case, achieving an optimal alignment may require making changes in the curriculum content or in the test content, or in both (see next principle).

References: Brophy (1983); Denham & Lieberman (1980); Doyle (1986).

3. Curricular alignment

> All components of the curriculum are aligned to create a cohesive programme for accomplishing instructional purposes and goals.

Research findings

Research indicates that educational policy-makers, textbook publishers and teachers often become so focused on content coverage or learning activities that they lose sight of the larger purposes and goals that are supposed to guide curriculum planning. Teachers typically plan by concentrating on the content they intend to cover and the steps involved in the activities their students will carry out, without giving much thought to the goals or intended outcomes of the instruction. Textbook publishers, in response to pressure from special interest groups, tend to keep expanding their content coverage. As a result, too many topics are covered in not enough depth; content exposition often lacks coherence and is cluttered with insertions; skills are taught separately from knowledge content rather than integrated with it; and in general, neither the students' texts nor the questions and activities suggested in the teachers' manuals are structured around powerful ideas connected to important goals.

Students taught using such textbooks may be asked to memorize parades of disconnected facts or to practise disconnected subskills in isolation instead of learning coherent networks of connected content structured around powerful ideas. These problems are often exacerbated by externally imposed assessment programmes that emphasize recognition of isolated bits of knowledge or performance of isolated subskills. Such problems can be minimized through goal-oriented curriculum development, in which curricular planning is guided by the overall purposes and goals of the instruction, not by miscellaneous content coverage pressures or test items.

In the classroom

A curriculum is not an end in itself; it is a means of helping students to learn what is considered essential for preparing them to fulfil adult roles in society and realize their potential as individuals. Its goals are learner outcomes–the knowledge, skills, attitudes, values and dispositions to action that society wishes to develop in its citizens. The goals are the reason for the existence of the curriculum, so that beliefs about what is needed to accomplish them should guide each step in curriculum planning and implementation. Goals are most likely to be attained if all of the curriculum's components (content clusters, instructional methods, learning activities and assessment tools) are selected because they are believed to be needed as means of helping students to accomplish the overall purposes and goals.

This involves planning curriculum and instruction to develop capabilities that students can use in their lives inside and outside school, both now and in the future. In this regard, it is important to emphasize goals of understanding, appreciation and life application. Understanding means that students learn both the individual elements in a network of related content and the connections among them, so that they can explain the content in their own words and connect it to their prior knowledge. Appreciation means that students value what they are learning because they understand that there are good reasons for learning it. Life application means that students retain their learning in a form that makes it usable when needed in other contexts.

Content developed with these goals in mind is likely to be retained as meaningful learning that is internally coherent, well connected with other meaningful learning and accessible for application. This is most likely to occur when the content itself is structured around powerful ideas and the development of this content through classroom lessons and learning activities focuses on these ideas and their connections.

References: Beck & McKeown (1988); Clark & Peterson (1986); Wang, Haertel & Walberg (1993).

4. Establishing learning orientations

Teachers can prepare students for learning by providing an initial structure to clarify intended outcomes and cue desired learning strategies.

Research findings

Research indicates the value of establishing a learning orientation by beginning lessons and activities with advance organizers or previews. These introductions facilitate students' learning by communicating the nature and purpose of the activity, connecting it to prior knowledge and cueing the kinds of student responses that the activity requires. This helps students to remain goal-oriented and strategic as they process information and respond to the questions or tasks embodied in the activity. Good lesson orientations also stimulate students' motivation to learn by communicating enthusiasm for the learning or helping students to appreciate its value or application potential.

In the classroom

Advance organizers orient students to what they will be learning before the instruction begins. They characterize the general nature of the activity and give students a structure within which to understand and connect the specifics that will be presented by the teacher or text. Such knowledge of the nature of the activity and the structure of its content helps students to focus on the main ideas and order their thoughts effectively. Therefore, before beginning any lesson or activity, the teacher should ensure that students know what they will be learning and why it is important for them to learn it.

Other ways to help students learn with a sense of purpose and direction include calling attention to the activity's goals, overviewing main ideas or major steps to be elaborated, pre-

tests that sensitize students to main points to learn, and pre-questions that stimulate their thinking about the topic.

References: Ausubel (1968); Brophy (1998); Meichenbaum & Biemiller (1998).

5. Coherent content

> To facilitate meaningful learning and retention, content is explained clearly and developed with emphasis on its structure and connections.

Research findings

Research indicates that networks of connected knowledge structured around powerful ideas can be learned with under-standing and retained in forms that make them accessible for application. In contrast, disconnected bits of information are likely to be learned only through low-level processes such as rote memorizing, and most of these bits either are soon for-gotten or are retained in ways that limit their accessibility. Similarly, skills are likely to be learned and used effectively if taught as strategies adapted to particular purposes and situa-tions, with attention to when and how to apply them; but students may not be able to integrate and use skills that are learned only by rote and practised only in isolation from the rest of the curriculum.

In the classroom

Whether in textbooks or in teacher-led instruction, information is easier to learn to the extent that it is coherent–the sequence of ideas or events makes sense and the relationships among them are apparent. Content is most likely to be organized coherently when it is selected in a principled way, guided by ideas about what students should learn from study-ing the topic.

When making presentations, providing explanations or giv-ing demonstrations, effective teachers project enthusiasm for the content and organize and sequence it so as to maximize its clarity and coherence. The teacher presents new information with reference to what students already know about the topic; proceeds in small steps sequenced in ways that are easy to fol-low; uses pacing, gestures and other oral communication skills

to support comprehension; avoids vague or ambiguous language and digressions that disrupt continuity; elicits students' responses regularly to stimulate active learning and ensure that each step is mastered before moving to the next; finishes with a review of main points, stressing general integrative concepts; and follows up with questions or assignments that require students to encode the material in their own words and apply or extend it to new contexts. If necessary, the teacher also helps students to follow the structure and flow of the content by using outlines or graphic organizers that depict relationships, study guides that call attention to key ideas, or task organizers that help students keep track of the steps involved and the strategies they use to complete these steps.

In combination, the principles calling for curricular alignment and for coherent content imply that, to enable students to construct meaningful knowledge that they can access and use in their lives outside school, teachers need to: (i) retreat from breadth of coverage in order to allow time to develop the most important content in greater depth; (ii) represent this important content as networks of connected information structured around powerful ideas; (iii) develop the content with a focus on explaining these important ideas and the connections among them; and (iv) follow up with authentic learning activities and assessment measures that provide students with opportunities to develop and display learning that reflects the intended outcomes of the instruction.

References: Beck & McKeown (1988); Good & Brophy (2000); Rosenshine (1968).

6. Thoughtful discourse

> Questions are planned to engage students in sustained discourse structured around powerful ideas.

Research findings

Besides presenting information and modelling application of skills, effective teachers structure a great deal of content-based discourse. They use questions to stimulate students to process and reflect on content, recognize relationships among and implications of its key ideas, think critically about it, and use it in problem solving, decision making or other higher-order applications. The discourse is not limited to rapidly paced recitation that elicits short answers to miscellaneous questions. Instead, it features sustained and thoughtful development of key ideas. Through participation in such discourse, students construct and communicate content-related understandings. In the process, they abandon naïve ideas or misconceptions and adopt the more sophisticated and valid ideas embedded in the instructional goals.

In the classroom

In the early stages of units when new content is introduced and developed, more time is spent in interactive lessons featuring teacher/student discourse than in independent work on assignments. The teacher plans sequences of questions designed to develop the content systematically and help students to construct understandings of it by relating it to their prior knowledge and collaborating in dialogue about it.

The forms and cognitive levels of these questions need to be suited to the instructional goals. Some primarily closed-end and factual questions might be appropriate when teachers are assessing prior knowledge or reviewing new learning, but accomplishing the most significant instructional goals requires open-ended questions that call for students to apply, analyse, synthesize or evaluate what they are learning. Some questions

will admit of a range of possible correct answers, and some will invite discussion or debate (e.g. concerning the relative merits of alternative suggestions for solving problems).

Because questions are intended to engage students in cognitive processing and construction of knowledge, they should ordinarily be addressed to the class as a whole. This encourages all students, not just the one eventually called on, to listen carefully and respond thoughtfully to each question. After posing a question, the teacher needs to pause to allow students enough time to process it and at least begin to formulate responses, especially if the question is complicated or requires students to engage in higher-order thinking.

Thoughtful discourse features sustained examination of a small number of related topics, in which students are invited to develop explanations, make predictions, debate alternative approaches to problems, or otherwise consider the content's implications or applications. The teacher presses students to clarify or justify their assertions, rather than accepting them indiscriminately. In addition to providing feedback, the teacher encourages students to explain or elaborate on their answers or to comment on classmates' answers. Frequently, discourse that begins in a question-and-answer format evolves into an exchange of views in which students respond to one another as well as to the teacher and respond to statements as well as to questions.

References: Good & Brophy (2000); Newmann (1990); Rowe (1986).

7. Practice and application activities

> Students need sufficient opportunities to practise and apply what they are learning, and to receive improvement-oriented feedback.

Research findings

There are three main ways in which teachers help their students to learn. First, they present information, explain concepts and model skills. Second, they ask questions and lead their students in discussion and other forms of discourse surrounding the content. Third, they engage students in activities or assignments that provide them with opportunities to practise or apply what they are learning. Research indicates that skills practised to a peak of smoothness and automaticity tend to be retained indefinitely, whereas skills that are mastered only partially tend to deteriorate. Most skills included in school curricula are learned best when practice is distributed across time and embedded within a variety of tasks. Thus, it is important to follow up thorough initial teaching with occasional review activities and with opportunities for students to use what they are learning in a variety of application contexts.

In the classroom

Practice is one of the most important yet least appreciated aspects of learning in classrooms. Little or no practice may be needed for simple behaviours such as pronouncing words, but practice becomes more important as learning becomes complex. Successful practice involves polishing skills that are already established at rudimentary levels in order to make them smoother, more efficient and more automatic, and not trying to establish such skills through trial and error.

Fill-in-the-blank worksheets, pages of mathematical computation problems and related tasks that engage students in memorizing facts or practising subskills in isolation from the

rest of the curriculum should be minimized. Instead, most practice should be embedded within application contexts that feature conceptual understanding of knowledge and self-regulated application of skills. Thus, most practice of reading skills is embedded within lessons involving reading and inter-preting extended text, most practice of writing skills is embed-ded within activities calling for authentic writing, and most practice of mathematics skills is embedded within problem-solving applications.

Opportunity to learn in school can be extended through homework assignments that are realistic in length and difficulty given the students' abilities to work independently. To ensure that students know what to do, the teacher can get them started on assignments in class, and then have them finish the work at home. An accountability system should be in place to ensure that students complete their homework assignments, and the work should be reviewed in class the next day.

To be useful, practice must involve opportunities not only to apply skills but also to receive timely feedback. Feedback should be informative rather than evaluative, helping students to assess their progress with respect to major goals and to understand and correct errors or misconceptions. At times when teachers are unable to circulate to monitor progress and provide feedback, they should arrange for students working on assignments to get feedback by consulting posted study guides or answer sheets or by asking peers designated to act as tutors or resource persons.

References: Brophy & Alleman (1991); Cooper (1994); Dempster (1991); Knapp (1995).

8. Scaffolding students' task engagement

> The teacher provides whatever assistance students need to enable them to engage in learning activities productively.

Research findings

Research on learning tasks suggests that activities and assignments should be sufficiently varied and interesting to motivate student engagement, sufficiently new or challenging to constitute meaningful learning experiences rather than needless repetition, and yet sufficiently easy to allow students to achieve high rates of success if they invest reasonable time and effort. The effectiveness of assignments is enhanced when teachers first explain the work and go over practice examples with students before releasing them to work independently, and then circulate to monitor progress and provide help when needed. The principle of teaching within the students' zones of proximal development implies that students will need explanation, modelling, coaching and other forms of assistance from their teachers, but also that this teacher structuring and scaffolding will be faded as the students' expertise develops. Eventually, students should become able to use what they are learning autonomously and to regulate their own productive task engagement.

In the classroom

Besides being well chosen, activities need to be effectively presented, monitored and followed up if they are to have their full impact. This means preparing students for an activity in advance, providing guidance and feedback during the activity, and leading the class in post-activity reflection afterwards. In introducing activities, teachers should stress their purposes in ways that will help students to engage in them with clear ideas about the goals to be accomplished. Then they might call stu-

dents' attention to relevant background knowledge, model strategies for responding to the task or scaffold by providing information about task requirements. If reading is involved, for example, teachers might summarize the main ideas, remind students about strategies for developing and monitoring their comprehension as they read (paraphrasing, summarizing, taking notes, asking themselves questions to check understanding), distribute study guides that call attention to key ideas and structural elements, or provide task organizers that help students to keep track of the steps involved and the strategies that they are using.

Once students begin working on activities or assignments, teachers should circulate to monitor their progress and provide assistance if necessary. Assuming that students have a general understanding of what to do and how to do it, these interventions can be kept brief and confined to minimal and indirect forms of help. If teacher assistance is too direct or extensive, teachers will end up carrying out tasks for students instead of helping them learn to carry out the tasks themselves.

Teachers also need to assess performance for completion and accuracy. When performance is poor, they will need to provide re-teaching and follow-up assignments designed to ensure that content is understood and skills are mastered.

Most assignments will not have their full effects unless they are followed by reflection or debriefing activities in which the teacher reviews the task with the students, provides general feedback about performance, and reinforces main ideas as they relate to overall goals. Reflection activities should also include opportunities for students to ask follow-up questions, share task-related observations or experiences, compare opinions, or in other ways deepen their appreciation of what they have learned and how it relates to their lives outside school.

References: Brophy & Alleman (1991); Rosenshine & Meister (1992); Shuell (1996); Tharp & Gallimore (1988).

9. Strategy teaching

> The teacher models and instructs students in learning and self-regulation strategies.

Research findings

General learning and study skills as well as domain-specific skills (such as constructing meaning from text, solving mathematical problems or reasoning scientifically) are most likely to be learned thoroughly and become accessible for application if they are taught as strategies to be brought to bear purposefully and implemented with metacognitive awareness and self-regulation. This requires comprehensive instruction that includes attention to propositional knowledge (what to do), procedural knowledge (how to do it) and conditional knowledge (when and why to do it). Strategy teaching is especially important for less able students who otherwise might not come to understand the value of consciously monitoring, self-regulating and reflecting upon their learning processes.

In the classroom

Many students do not develop effective learning and problem-solving strategies on their own but can acquire them through modelling and explicit instruction from their teachers. Poor readers, for example, can be taught reading comprehension strategies such as keeping the purpose of an assignment in mind when reading; activating relevant background knowledge; identifying major points in attending to the outline and flow of content; monitoring understanding by generating and trying to answer questions about the content; or drawing and testing inferences by making interpretations, predictions and conclusions. Instruction should include not only demonstrations of and opportunities to apply the skill itself but also explanations of the purpose of the skill (what it does for the learner) and the occasions on which it would be used.

Strategy teaching is likely to be most effective when it includes cognitive modelling: the teacher thinks out loud while modelling use of the strategy. Cognitive modelling makes overt

the otherwise covert thought processes that guide use of the strategy in a variety of contexts. It provides learners with first-person language ('self talk') that they can adapt directly when using the strategy themselves. This eliminates the need for translation that is created when instruction is presented in the impersonal third-person language of explanation or even the second-person language of coaching.

In addition to strategies used in particular domains or types of assignments, teachers can model and instruct their students in general study skills and learning strategies such as rehearsal (repeating material to remember it more effectively), elaboration (putting material into one's own words and relating it to prior knowledge), organization (outlining material to highlight its structure and remember it), comprehension monitoring (keeping track of the strategies used to construct understandings and the degree of success achieved with them, and adjusting strategies accordingly), and affect monitoring (maintaining concentration and task focus, and minimizing performance anxiety and fear of failure).

When providing feedback as students work on assignments and when leading subsequent reflection activities, teachers can ask questions or make comments that help students to monitor and reflect on their learning. Such monitoring and reflection should focus not only on the content being learned, but also on the strategies that the students are using to process the content and solve problems. This will help the students to refine their strategies and regulate their learning more systematically.

References: Meichenbaum & Biemiller (1998); Pressley & Beard El-Dinary (1993); Weinstein & Mayer (1986).

10. Co-operative learning

> Students often benefit from working in pairs or small groups to construct understandings or help one another master skills.

Research findings

Research indicates that there is often much to be gained by arranging for students to collaborate in pairs or small groups as they work on activities and assignments. Co-operative learning promotes affective and social benefits such as increased student interest in and valuing of subject matter, and increases in positive attitudes and social interactions among students who differ in gender, race, ethnicity, achievement levels and other characteristics.

Co-operative learning also creates the potential for cognitive and metacognitive benefits by engaging students in discourse that requires them to make their task-related information-processing and problem-solving strategies explicit (and thus available for discussion and reflection). Students are likely to show improved achievement outcomes when they engage in certain forms of co-operative learning as an alternative to completing assignments on their own.

In the classroom

Traditional approaches to instruction feature whole-class lessons followed by independent seatwork time during which students work alone (and usually silently) on assignments. Co-operative learning approaches retain the whole-class lessons but replace part of the individual seatwork time with opportunities for students to work together in pairs or small groups on follow-up practice and application activities. Co-operative learning can be used with activities ranging from drill and practice to learning facts and concepts, discussion and problem solving. It is perhaps most valuable as a way of engaging students in meaningful learning with authentic tasks in a social setting. Students have more chances to talk in pairs or small

groups than in whole-class activities, and shy students are more likely to feel comfortable expressing ideas in these more intimate settings.

Some forms of co-operative learning call for students to help one another achieve individual learning goals, for example by discussing how to respond to assignments, checking work, or providing feedback or tutorial assistance. Other forms of co-operative learning call for students to work together to achieve a group goal by pooling their resources and sharing the work. For example, the group might conduct an experiment, assemble a collage, or prepare a research report to be presented to the rest of the class. Co-operative learning models that call for students to work together to produce a group product often feature a division of labour among group participants (e.g. to prepare a biographical report, one group member will assume responsibility for studying the person's early life, another for the person's major accomplishments, another for the person's effects on society, and so on).

Co-operative learning methods are most likely to enhance learning outcomes if they combine group goals with individual accountability. That is, each group member will be held accountable for accomplishing the activity's learning goals (students know that any member of the group may be called on to answer any one of the group's questions or that they will all be tested individually on what they are learning).

Activities used in co-operative learning formats should be well suited to those formats. Some activities are most naturally carried out by individuals working alone, others by students working in pairs, and still others by small groups of three to six students.

Students should receive whatever instruction and scaffolding they may need to prepare them for productive engagement in co-operative learning activities. For example, teachers may need to show their students how to listen, share, integrate the ideas of others and handle disagreements constructively. During times when students are working in pairs or small groups, the teacher should circulate to monitor progress, make sure that groups are working productively and provide any assistance needed.

References: Bennett & Dunne (1992); Johnson & Johnson (1994); Slavin (1990).

11. Goal-oriented assessment

> The teacher uses a variety of formal and informal assessment methods to monitor progress towards learning goals.

Research findings

A well-developed curriculum includes strong and functional assessment components. These assessment components are aligned with the curriculum's goals, and so they are integrated with its content, instructional methods and learning activities, and designed to evaluate progress towards its major intended outcomes.

Comprehensive assessment does not just document students' ability to supply acceptable answers to questions or problems; it also examines the students' reasoning and problem-solving processes. Effective teachers routinely monitor their students' progress in this fashion, using both formal tests or performance evaluations and informal assessments of students' contributions to lessons and work on assignments.

In the classroom

Effective teachers use assessment for evaluating students' progress in learning and for planning curriculum improvements, not just for generating grades. Good assessment includes data from many sources besides paper-and-pencil tests, and it addresses the full range of goals or intended outcomes (not only knowledge but also higher-order thinking skills and content-related values and dispositions). Standardized, norm-referenced tests might comprise part of the assessment programme (these tests are useful to the extent that they measure intended outcomes of the curriculum and attention is paid to students' performance on each individual item, not just total scores). However, standardized tests should ordinarily be supplemented with publisher-supplied curriculum-embedded tests (when these appear useful) and with teacher-made tests that focus on learning goals that are emphasized in instruction but not in external testing sources.

In addition, learning activities and sources of data other than tests should be used for assessment purposes. Everyday lessons and activities provide opportunities to monitor the progress of the class as a whole and of individual students, and tests can be augmented with performance evaluations such as laboratory tasks and observation checklists, portfolios of student papers or projects, and essays or other assignments that call for higher-order thinking and application. A broad view of assessment helps to ensure that the assessment component includes authentic activities that provide students with opportunities to synthesize and reflect on what they are learning, think critically and creatively about it, and apply it in problem-solving and decision-making contexts.

In general, assessment should be treated as an ongoing and integral part of each instructional unit. Results should be scrutinized to identify learner needs, misunderstandings or misconceptions that may need attention; to suggest potential adjustment in curriculum goals, instructional materials or teaching plans; and to detect weaknesses in the assessment practices themselves.

References: Dempster (1991); Stiggins (1997); Wiggins (1993).

12. Achievement expectations

> The teacher establishes and follows through on appropriate expectations for learning outcomes.

Research findings

Research indicates that effective schools feature strong academic leadership that produces consensus on goal priorities and commitment to instructional excellence, as well as positive teacher attitudes towards students and expectations regarding their abilities to master the curriculum. Teacher effects research indicates that teachers who elicit strong achievement gains accept responsibility for doing so. They believe that their students are capable of learning and that they (the teachers) are capable of and responsible for teaching them successfully. If students do not learn something the first time, they teach it again, and if the regular curriculum materials do not do the job, they find or develop others that will.

In the classroom

Teachers' expectations concerning what their students are capable of accomplishing (with teacher help) tend to shape both what teachers attempt to elicit from their students and what the students come to expect from themselves. Thus, teachers should form and project expectations that are as positive as they can be while still remaining realistic. Such expectations should represent genuine beliefs about what can be achieved and therefore should be taken seriously as goals towards which to work in instructing students.

It is helpful if teachers set goals for the class and for individuals in terms of floors (minimally acceptable standards), not ceilings. Then they can let group progress rates, rather than limits adopted arbitrarily in advance, determine how far the class can go within the time available. They can keep their expectations for individual students current by monitoring their progress closely and by stressing current performance over past history.

At the very least, teachers should expect all their students to progress sufficiently to enable them to perform satisfactorily at the next level. This implies holding all students accountable for participating in lessons and learning activities and for turning in careful and completed work on assignments. It also implies that, in addition to the other elements of good teaching summarized in the preceding principles, struggling students will receive whatever extra time, instruction and encouragement are needed to enable them to meet expectations.

When individualizing instruction and giving students feedback, teachers should emphasize continuous progress relative to previous levels of mastery rather than how students compare with their classmates or with standardized test norms. Instead of merely evaluating relative levels of success, teachers can diagnose learning difficulties and provide feedback accordingly. If students have not understood an explanation or demonstration, teachers can follow through by re-teaching (if necessary, in a different way rather than by merely repeating the original instruction).

In general, teachers are likely to be most successful when they think in terms of stretching students' minds by stimulating them and encouraging them to achieve as much as they can, not in terms of 'protecting' them from failure or embarrassment.

References: Brophy (1998); Creemers & Scheerens (1989); Good & Brophy (2000); Shuell (1996); Teddlie & Stringfield (1993).

Conclusion

To date, most research on teaching has been conducted in the United States, Canada, Western Europe and Australia, and so the degree to which findings apply to other countries has yet to be addressed. The principles presented in this booklet are believed to apply universally, however, for two reasons. First, research done all over the world suggests that schooling is much more similar than different across countries and cultures. The day is divided into periods used for teaching each of the subjects included in the curriculum, and teaching includes whole-class lessons in which content is developed through teacher explanation and teacher/student interaction, followed by practice and application activities that students work on individually or in pairs or small groups. Second, the principles refer to generic aspects of teaching that cut across grade levels and school subjects, not to particular curriculum content. In summary, these principles ought to apply universally because they focus on basic and universal aspects of formal schooling. They still require adaptation to the local context, however, including relevant characteristics of the nation's school system and the students' cultures.

The generic principles featured in this booklet need to be supplemented with more specific principles that apply to the teaching of particular school subjects to particular types of students. Readers interested in planning instruction for particular grade levels and subject areas can consult the scholarly literature in the subject areas for elaborations on and additions to the principles outlined here.

Finally, although twelve principles are highlighted for emphasis and discussed individually, each principle should be applied in conjunction with the others. That is, the principles are meant to be understood as mutually supportive components of a coherent approach to teaching in which the teacher's plans and expectations, the classroom learning environment and management system, the curriculum content and instructional materials, and the learning activities and assessment methods are all aligned as means of helping students attain intended outcomes.

References

Ausubel, D. 1968. *Educational psychology: a cognitive view.* New York, Holt, Rinehart & Winston.

Beck, I.; McKeown, M. 1988. Toward meaningful accounts in history texts for young learners. *Educational researcher* (Washington, DC), vol. 17, n°. 6, p. 31-39.

Bennett, N.; Dunne, E. 1992. *Managing small groups.* New York, Simon & Schuster.

Brophy, J. 1983. Classroom organization and management. *The elementary school journal* (Chicago, IL), vol. 83, p. 265-85.

—.1998. *Motivating students to learn.* Boston, McGraw-Hill.

Brophy, J.; Alleman, J. 1991. Activities as instructional tools: a framework for analysis and evaluation. *Educational researcher* (Washington, DC), vol. 20, n°. 4, p. 9-23.

Clark, C.; Peterson, P. 1986. Teachers' thought processes. *In:* Wittrock, M.C., ed. *Handbook of research on teaching*, 3rd ed., p. 225-296. New York, Macmillan.

Cooper, H. 1994. *The battle over homework: an administrator's guide to setting sound and effective policies.* Thousand Oaks, CA, Corwin.

Creemers, B.; Scheerens, J., guest eds. 1989. Developments in school effectiveness research. *International journal of educational research* (Oxford, UK), vol. 13, p. 685-825.

Dempster, F. 1991. Synthesis of research on reviews and tests. *Educational leadership* (Alexandria, VA), vol. 48, n°. 7, p. 71–76.

Denham, C.; Lieberman, A., eds. 1980. *Time to learn.* Washington, DC, National Institute of Education.

Doyle, W. 1986. Classroom organization and management. *In:* Wittrock, M.C., ed. *Handbook of research on teaching*, 3rd ed., p. 392-431. New York, Macmillan.

Good, T.; Brophy, J. 1986. School effects. *In:* Wittrock, M.C., ed. *Handbook of research on teaching*, 3rd ed., p. 570-602. New York, Macmillan.

—; — . 2000. *Looking in classrooms*, 8th ed. New York, Longman.

Johnson, D.; Johnson, R. 1994. *Learning together and alone: cooperative, competitive, and individualistic learning*, 4th ed. Boston, Allyn & Bacon.

Knapp, M. 1995. *Teaching for meaning in high-poverty classrooms.* New York, Teachers College Press.

Meichenbaum, D.; Biemiller, A. 1998. *Nurturing independent learners: helping students take charge of their learning.* Cambridge, MA, Brookline.

Newmann, F. 1990. Qualities of thoughtful social studies classes: an empirical profile. *Journal of curriculum studies* (Basingstoke, UK), vol. 22, p. 253-275.

Pressley, M.; Beard El-Dinary, P., guest eds. 1993. Special issue on strategies instruction. *The elementary school journal* (Chicago, IL), vol. 94, p. 105-284.

Rosenshine, B. 1968. To explain: a review of research. *Educational leadership* (Alexandria, VA), n°. 26, p. 275-280.

Rosenshine, B.; Meister, C. 1992. The use of scaffolds for teaching higher-level cognitive strategies. *Educational leadership* (Alexandria, VA), vol. 49, n°. 7, p. 26-33.

Rowe, M. 1986. Wait time: slowing down may be a way of speeding up! *Journal of teacher education* (Thousand Oaks, CA), vol. 37, p. 43-50.

Sergiovanni, T. 1994. *Building community in schools.* San Francisco, Jossey-Bass.

Shuell, T. 1996. Teaching and learning in a classroom context. *In:* Berliner, D.; Calfee, R., eds. *Handbook of educational psychology,* p. 726-764. New York, Macmillan.

Slavin, R. 1990. *Cooperative learning: theory, research, and practice.* Englewood Cliffs, NJ, Prentice-Hall.

Stiggins, R. 1997. *Student-centered classroom assessment,* 2nd ed. Upper Saddle River, NJ, Prentice-Hall.

Teddlie, C.; Stringfield, S. 1993. *Schools make a difference: lessons learned from a 10–year study of school effects.* New York, Teachers College Press.

Tharp, R.; Gallimore, R. 1988. *Rousing minds to life: teaching, learning, and schooling in social context.* Cambridge, Cambridge University Press.

Wang, M.; Haertel, G.; Walberg, H. 1993. Toward a knowledge base for school learning. *Review of educational research* (Washington, DC), vol. 63, p. 249-294.

Weinstein, C.; Mayer, R. 1986. The teaching of learning strategies. *In:* Wittrock, M.C., ed. *Handbook of research on teaching,* 3rd ed., p. 315–27. New York, Macmillan.

Wiggins, G. 1993. *Assessing student performance: exploring the purpose and limits of testing.* San Francisco, Jossey-Bass.

The International Academy of Education

The International Academy of Education (IAE) is a not-for-profit scientific association that promotes educational research, its dissemination, and the implementation of its implications. Founded in 1986, the Academy is dedicated to strengthening the contributions of research, solving critical educational problems throughout the world, and providing better communication among policy makers, researchers and practitioners. The seat of the Academy is at the Royal Academy of Science, Literature and Arts in Brussels, Belgium, and its co-ordinating centre is at Curtin University of Technology in Perth, Australia.

The general aim of the IAE is to foster scholarly excellence in all fields of education. Towards this end, the Academy provides timely syntheses of research-based evidence of international importance. The Academy also provides critiques of research, its evidentiary basis, and its application to policy.

The current members of the Board of Directors of the Academy are:

Principles of Instruction

Research-Based Strategies That All Teachers Should Know

By Barak Rosenshine

This article presents 10 research-based principles of instruction, along with suggestions for classroom practice. These principles come from three sources: (a) research in cognitive science, (b) research on master teachers, and (c) research on cognitive supports. Each is briefly explained below.

A: Research in cognitive science: This research focuses on how our brains acquire and use information. This cognitive research also provides suggestions on how we might overcome the limitations of our working memory (i.e., the mental "space" in which thinking occurs) when learning new material.

B: Research on the classroom practices of master teachers: Master teachers are those teachers whose classrooms made the highest gains on achievement tests. In a series of studies, a wide range of teachers were observed as they taught, and the investigators coded how they presented new material, how and whether they checked for student understanding, the types of support they provided to their students, and a number of other instructional activities. By also gathering student achievement data, researchers were able to identify the ways in which the more and less effective teachers differed.

C: Research on cognitive supports to help students learn complex tasks: Effective instructional procedures—such as thinking aloud, providing students with scaffolds, and providing students with models—come from this research.

Barak Rosenshine is an emeritus professor of educational psychology in the College of Education at the University of Illinois at Urbana-Champaign. A distinguished researcher, he has spent much of the past four decades identifying the hallmarks of effective teaching. He began his career as a high school history teacher in the Chicago public schools. This article is adapted with permission from Principles of Instruction *by Barak Rosenshine. Published by the International Academy of Education in 2010, the original report is available at www.ibe.unesco.org/fileadmin/user_upload/Publications/Educational_Practices/EdPractices_21.pdf.*

Even though these are three very different bodies of research, there is *no conflict at all* between the instructional suggestions that come from each of these three sources. In other words, these three sources supplement and complement each other. The fact that the instructional ideas from three different sources supplement and complement each other gives us faith in the validity of these findings.

Education involves helping a novice develop strong, readily accessible background knowledge. It's important that background knowledge be readily accessible, and this occurs when knowledge is well rehearsed and tied to other knowledge. The most effective teachers ensured that their students efficiently acquired, rehearsed, and connected background knowledge by providing a good deal of instructional support. They provided this support by teaching new material in manageable amounts, modeling, guiding student practice, helping students when they made errors, and providing for sufficient practice and review. Many of these teachers also went on to experiential, hands-on activities, but they always did the experiential activities *after*, not before, the basic material was learned.

The following is a list of some of the instructional principles that have come from these three sources. These ideas will be described and discussed in this article:

- Begin a lesson with a short review of previous learning.[1]
- Present new material in small steps with student practice after each step.[2]
- Ask a large number of questions and check the responses of all students.[3]
- Provide models.[4]
- Guide student practice.[5]
- Check for student understanding.[6]
- Obtain a high success rate.[7]
- Provide scaffolds for difficult tasks.[8]
- Require and monitor independent practice.[9]
- Engage students in weekly and monthly review.[10]

1. Begin a lesson with a short review of previous learning: Daily review can strengthen previous learning and can lead to fluent recall.

Research findings

Daily review is an important component of instruction. Review can help us strengthen the connections among the material we have learned. The review of previous learning can help us recall words, concepts, and procedures effortlessly and automatically when we need this material to solve problems or to understand new material. The development of expertise requires thousands of hours of practice, and daily review is one component of this practice.

For example, daily review was part of a successful experiment in elementary school mathematics. Teachers in the experiment were taught to spend eight minutes every day on review. Teachers used this time to check the homework, go over problems where there were errors, and practice the concepts and skills that needed to become automatic. As a result, students in these classrooms had higher achievement scores than did students in other classrooms.

Daily practice of vocabulary can lead to seeing each practiced word as a unit (i.e., seeing the whole word automatically rather than as individual letters that have to be sounded out and blended). When students see words as units, they have more space available in their working memory, and this space can now be used for comprehension. Mathematical problem solving is also improved when the basic skills (addition, multiplication, etc.) are overlearned and become automatic, thus freeing working-memory capacity.

In the classroom

The most effective teachers in the studies of classroom instruction understood the importance of practice, and they began their lessons with a five- to eight-minute review of previously covered material. Some teachers reviewed vocabulary, formulae, events, or previously learned concepts. These teachers provided additional practice on facts and skills that were needed for recall to become automatic.

Effective teacher activities also included reviewing the concepts and skills that were necessary to do the homework, having students correct each others' papers, and asking about points on which the students had difficulty or made errors. These reviews

ensured that the students had a firm grasp of the skills and concepts that would be needed for the day's lesson.

Effective teachers also reviewed the knowledge and concepts that were relevant for that day's lesson. It is important for a teacher to help students recall the concepts and vocabulary that will be relevant for the day's lesson because our working memory is very limited. If we do not review previous learning, then we will have to make a special effort to recall old material while learning new material, and this makes it difficult for us to learn the new material.

Daily review is particularly important for teaching material that will be used in subsequent learning. Examples include reading sight words (i.e., any word that is known by a reader automatically), grammar, math facts, math computation, math factoring, and chemical equations.

When planning for review, teachers might want to consider which words, math facts, procedures, and concepts need to

> The most effective teachers ensured that students efficiently acquired, rehearsed, and connected knowledge. Many went on to hands-on activities, but always *after*, not before, the basic material was learned.

become automatic, and which words, vocabulary, or ideas need to be reviewed before the lesson begins.

In addition, teachers might consider doing the following during their daily review:

- Correct homework.
- Review the concepts and skills that were practiced as part of the homework.
- Ask students about points where they had difficulties or made errors.
- Review material where errors were made.
- Review material that needs overlearning (i.e., newly acquired skills should be practiced well beyond the point of initial mastery, leading to automaticity).

2. Present new material in small steps with student practice after each step: Only present small amounts of new material at any time, and then assist students as they practice this material.

Research findings

Our working memory, the place where we process information, is small. It can only handle a few bits of information at once—too much information swamps our working memory. Presenting too much material at once may confuse students because their working memory will be unable to process it.

Therefore, the more effective teachers do not overwhelm their students by presenting too much new material at once. Rather,

these teachers only present small amounts of new material at any time, and then assist the students as they practice this material. Only after the students have mastered the first step do teachers proceed to the next step.

The procedure of first teaching in small steps and then guiding student practice represents an appropriate way of dealing with the limitation of our working memory.

In the classroom
The more successful teachers did not overwhelm their students by presenting too much new material at once. Rather, they presented only small amounts of new material at one time, and they
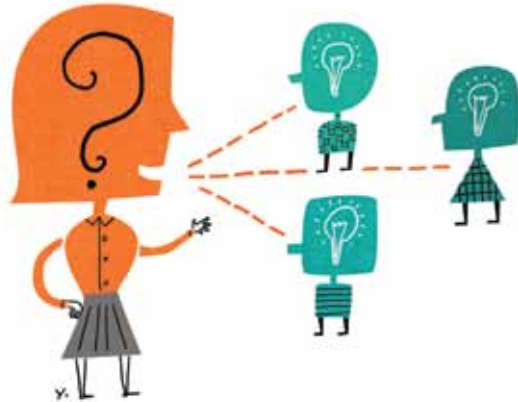
taught in such a way that each point was mastered before the next point was introduced. They checked their students' understanding on each point and retaught material when necessary.

Some successful teachers taught by giving a series of short presentations using many examples. The examples provided concrete learning and elaboration that were useful for processing new material.

Teaching in small steps requires time, and the more effective teachers spent more time presenting new material and guiding student practice than did the less effective teachers. In a study of mathematics instruction, for instance, the most effective mathematics teachers spent about 23 minutes of a 40-minute period in lecture, demonstration, questioning, and working examples. In contrast, the least effective teachers spent only 11 minutes presenting new material. The more effective teachers used this extra time to provide additional explanations, give many examples, check for student understanding, and provide sufficient instruction so that the students could learn to work independently without difficulty. In one study, the least effective teachers asked only nine questions in a 40-minute period. Compared with the successful teachers, the less effective teachers gave much shorter presentations and explanations, and then passed out worksheets and told students to solve the problems. The less successful teachers were then observed going from student to student and having to explain the material again.

Similarly, when students were taught a strategy for summarizing a paragraph, an effective teacher taught the strategy using small steps. First, the teacher modeled and thought aloud as she identified the topic of a paragraph. Then, she led practice on iden-

tifying the topics of new paragraphs. Then, she taught students to identify the main idea of a paragraph. The teacher modeled this step and then supervised the students as they practiced both finding the topic and locating the main idea. Following this, the teacher taught the students to identify the supporting details in a paragraph. The teacher modeled and thought aloud, and then the students practiced. Finally, the students practiced carrying out all three steps of this strategy. Thus, the strategy of summarizing a paragraph was divided into smaller steps, and there was modeling and practice at each step.

3. Ask a large number of questions and check the responses of all students: Questions help students practice new information and connect new material to their prior learning.

Research findings
Students need to practice new material. The teacher's questions and student discussion are a major way of providing this necessary practice. The most successful teachers in these studies spent more than half of the class time lecturing, demonstrating, and asking questions.

Questions allow a teacher to determine how well the material has been learned and whether there is a need for additional instruction. The most effective teachers also ask students to explain the process they used to answer the question, to explain how the answer was found. Less successful teachers ask fewer questions and almost no process questions.

In the classroom
In one classroom-based experimental study, one group of teachers was taught to follow the presentation of new material with lots of questions.[11] They were taught to increase the number of factual questions and process questions they asked during this guided practice. Test results showed that their students achieved higher scores than did students whose teachers did not receive the training.

Imaginative teachers have found ways to involve all students in answering questions. Examples include having all students:

- Tell the answer to a neighbor.
- Summarize the main idea in one or two sentences, writing the summary on a piece of paper and sharing this with a neighbor, or repeating the procedures to a neighbor.
- Write the answer on a card and then hold it up.
- Raise their hands if they know the answer (thereby allowing the teacher to check the entire class).
- Raise their hands if they agree with the answer that someone else has given.

Across the classrooms that researchers observed, the purpose of all these procedures was to provide active participation for the students and also to allow the teacher to see how many students were correct and confident. The teacher may then reteach some material when it was considered necessary. An alternative was for students to write their answers and then trade papers with each other.

Other teachers used choral responses to provide sufficient practice when teaching new vocabulary or lists of items. This made the practice seem more like a game. To be effective, how-

ever, all students needed to start together, on a signal. When students did not start together, only the faster students answered.

In addition to asking questions, the more effective teachers facilitated their students' rehearsal by providing explanations, giving more examples, and supervising students as they practiced the new material.

The following is a series of stems[12] for questions that teachers might ask when teaching literature, social science content, or science content to their students. Sometimes, students may also develop questions from these stems to ask questions of each other.
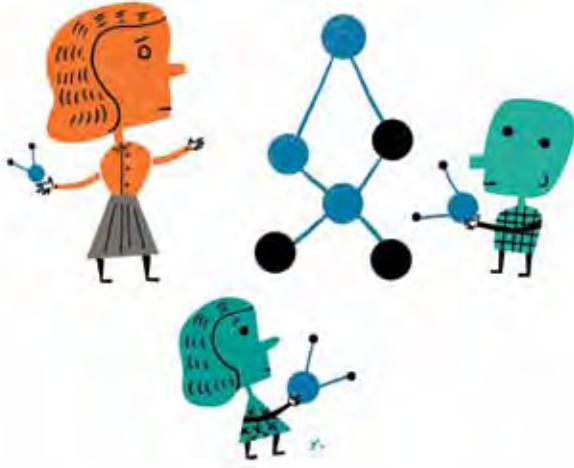
How are _____ and _____ alike?
What is the main idea of _____?
What are the strengths and weaknesses of _____?
In what way is _____ related to _____?
Compare _____ and _____ with regard to _____.
What do you think causes _____?
How does _____ tie in with what we have learned before?
Which one is the best _____, and why?
What are some possible solutions for the problem of _____?
Do you agree or disagree with this statement: _____?
What do you still not understand about _____?

4. Provide models: Providing students with models and worked examples can help them learn to solve problems faster.

Research findings
Students need cognitive support to help them learn to solve problems. The teacher modeling and thinking aloud while demonstrating how to solve a problem are examples of effective cognitive support. Worked examples (such as a math problem for which the teacher not only has provided the solution but has clearly laid out each step) are another form of modeling that has been developed by researchers. Worked examples allow students to focus on the specific steps to solve problems and thus reduce the cognitive load on their working memory. Modeling and worked examples have been used successfully in mathematics, science, writing, and reading comprehension.

In the classroom
Many of the skills that are taught in classrooms can be conveyed by providing prompts, modeling use of the prompt, and then guid-

ing students as they develop independence. When teaching reading comprehension strategies, for example, effective teachers provided students with prompts that the students could use to ask themselves questions about a short passage. In one class, students were given words such as "who," "where," "why," and "how" to help them begin a question. Then, everyone read a passage and the teacher modeled how to use these words to ask questions. Many examples were given.

Next, during guided practice, the teacher helped the students practice asking questions by helping them select a prompt and

> Many of the skills taught in classrooms can be conveyed by providing prompts, modeling use of the prompt, and then guiding students as they develop independence.

develop a question that began with that prompt. The students practiced this step many times with lots of support from the teacher.

Then, the students read new passages and practiced asking questions on their own, with support from the teacher when needed. Finally, students were given short passages followed by questions, and the teacher expressed an opinion about the quality of the students' questions.

This same procedure—providing a prompt, modeling, guiding practice, and supervising independent practice—can be used for many tasks. When teaching students to write an essay, for example, an effective teacher first modeled how to write each paragraph, then the students and teacher worked together on two or more new essays, and finally students worked on their own with supervision from the teacher.

Worked examples are another form of modeling that has been used to help students learn how to solve problems in mathematics and science. A worked example is a step-by-step demonstration of how to perform a task or how to solve a problem. The presentation of worked examples begins with the teacher modeling and explaining the steps that can be taken to solve a specific problem. The teacher also identifies and explains the underlying principles for these steps.

Usually, students are then given a series of problems to complete at their desks as independent practice. But, in research carried out in Australia, students were given a mixture of problems to solve and worked examples. So, during independent practice, students first studied a worked example, then they solved a problem; then they studied another worked example and solved another problem. In this way, the worked examples showed students how to focus on the essential parts of the problems. Of course, not all students studied the worked examples. To correct

this problem, the Australian researchers also presented partially completed problems in which students had to complete the missing steps and thus pay more attention to the worked example.

5. Guide student practice: Successful teachers spend more time guiding students' practice of new material.

Research findings
It is not enough simply to present students with new material, because the material will be forgotten unless there is sufficient rehearsal. An important finding from information-processing research is that students need to spend additional time rephrasing, elaborating, and summarizing new material in order to store this material in their long-term memory. When there has been sufficient rehearsal, the students are able to retrieve this material

easily and thus are able to make use of this material to foster new learning and aid in problem solving. But when the rehearsal time is too short, students are less able to store, remember, or use the material. As we know, it is relatively easy to place something in a filing cabinet, but it can be very difficult to recall where exactly we filed it. Rehearsal helps us remember where we filed it so we can access it with ease when needed.

A teacher can facilitate this rehearsal process by asking questions; good questions require students to process and rehearse the material. Rehearsal is also enhanced when students are asked to summarize the main points, and when they are supervised as they practice new steps in a skill. The quality of storage in long-term memory will be weak if students only skim the material and do not engage in it. It is also important that all students process the new material and receive feedback, so they do not inadvertently store partial information or a misconception in long-term memory.

In the classroom
In one study, the more successful teachers of mathematics spent more time presenting new material and guiding practice. The more successful teachers used this extra time to provide additional explanations, give many examples, check for student understanding, and provide sufficient instruction so that the students could learn to work independently without difficulty. In contrast, the less successful teachers gave much shorter presentations and explanations, and then they passed out worksheets and told stu-

dents to work on the problems. Under these conditions, the students made too many errors and had to be retaught the lesson.

The most successful teachers presented only small amounts of material at a time. After this short presentation, these teachers then guided student practice. This guidance often consisted of the teacher working the first problems at the blackboard and explaining the reason for each step, which served as a model for the students. The guidance also included asking students to come to the blackboard to work out problems and discuss their procedures. Through this process, the students seated in the classroom saw additional models.

Although most teachers provided some guided practice, the most successful teachers spent more time in guided practice, more time asking questions, more time checking for understanding, more time correcting errors, and more time having students work out problems with teacher guidance.

Teachers who spent more time in guided practice and had higher success rates also had students who were more engaged during individual work at their desks. This finding suggests that, when teachers provided sufficient instruction during guided practice, the students were better prepared for the independent practice (e.g., seatwork and homework activities), but when the guided practice was too short, the students were not prepared for the seatwork and made more errors during independent practice.

6. Check for student understanding: Checking for student understanding at each point can help students learn the material with fewer errors.

Research findings
The more effective teachers frequently checked to see if all the students were learning the new material. These checks provided some of the processing needed to move new learning into long-term memory. These checks also let teachers know if students were developing misconceptions.

In the classroom
Effective teachers also stopped to check for student understanding. They checked for understanding by asking questions, by asking students to summarize the presentation up to that point or to repeat directions or procedures, or by asking students whether they agreed or disagreed with other students' answers. This checking has two purposes: (a) answering the questions might cause the students to elaborate on the material they have learned and augment connections to other learning in their long-term memory, and (b) alerting the teacher to when parts of the material need to be retaught.

In contrast, the less effective teachers simply asked, "Are there any questions?" and, if there were no questions, they assumed the students had learned the material and proceeded to pass out worksheets for students to complete on their own.

Another way to check for understanding is to ask students to think aloud as they work to solve mathematical problems, plan an essay, or identify the main idea in a paragraph. Yet another check is to ask students to explain or defend their position to others. Having to explain a position may help students integrate and elaborate their knowledge in new ways, or may help identify gaps in their understanding.

Another reason for the importance of teaching in small steps, guiding practice, and checking for understanding (as well as obtaining a high success rate, which we'll explore in principle 7) comes from the fact that we all construct and reconstruct knowledge as we learn and use what we have learned. We cannot simply repeat what we hear word for word. Rather, we connect our understanding of the new information to our existing concepts or "schema," and we then construct a mental summary (i.e., the gist of what we have heard). However, when left on their own, many students make errors in the process of constructing this mental summary. These errors occur, particularly, when the information is new and the student does not have adequate or well-formed background knowledge. These constructions are not errors so much as attempts by the students to be logical in an area where their background knowledge is weak. These errors are so common that there is a research literature on the development and correc-

vised student practice), and by giving sufficient practice on each part before proceeding to the next step. These teachers frequently checked for understanding and required responses from all students.

It is important that students achieve a high success rate during instruction and on their practice activities. Practice, we are told, makes perfect, but practice can be a disaster if students are practicing errors! If the practice does not have a high success level, there is a chance that students are practicing and learning errors. Once errors have been learned, they are very difficult to overcome.

As discussed in the previous section, when we learn new material, we construct a gist of this material in our long-term memory. However, many students make errors in the process of constructing this mental summary. These errors can occur when the information is new and the student did not have adequate or

The most successful teachers spent more time in guided practice, more time asking questions, more time checking for understanding, and more time correcting errors.

tion of student misconceptions in science. Providing guided practice after teaching small amounts of new material, and checking for student understanding, can help limit the development of misconceptions.

7. Obtain a high success rate: It is important for students to achieve a high success rate during classroom instruction.

Research findings
In two of the major studies on the impact of teachers, the investigators found that students in classrooms with more effective teachers had a higher success rate, as judged by the quality of their oral responses during guided practice and their individual work. In a study of fourth-grade mathematics, it was found that 82 percent of students' answers were correct in the classrooms of the most successful teachers, but the least successful teachers had a success rate of only 73 percent. A high success rate during guided practice also leads to a higher success rate when students are working on problems on their own.

The research also suggests that the optimal success rate for fostering student achievement appears to be about 80 percent. A success rate of 80 percent shows that students are learning the material, and it also shows that the students are challenged.

In the classroom
The most effective teachers obtained this success level by teaching in small steps (i.e., by combining short presentations with super-

well-formed background knowledge. These constructions are not errors so much as attempts by the students to be logical in an area where their background knowledge is weak. But students are more likely to develop misconceptions if too much material is presented at once, and if teachers do not check for student understanding. Providing guided practice after teaching small amounts of new material, and checking for student understanding, can help limit the development of misconceptions.

I once observed a class where an effective teacher was going from desk to desk during independent practice and suddenly realized that the students were having difficulty. She stopped the work, told the students not to do the problems for homework, and said she would reteach this material the next day. She stopped the work because she did not want the students to practice errors.

Unless all students have mastered the first set of lessons, there is a danger that the slower students will fall further behind when the next set of lessons is taught. So there is a need for a high success rate for *all* students. "Mastery learning" is a form of instruction where lessons are organized into short units and all students are required to master one set of lessons before they proceed to the next set. In mastery learning, tutoring by other students or by teachers is provided to help students master each unit. Variations of this approach, particularly the tutoring, might be useful in many classroom settings.

8. Provide scaffolds for difficult tasks: The teacher provides students with temporary supports and scaffolds to assist them when they learn difficult tasks.

Research findings
Investigators have successfully provided students with scaffolds, or instructional supports, to help them learn difficult tasks. A scaffold is a temporary support that is used to assist a learner. These scaffolds are gradually withdrawn as learners become more competent, although students may continue to rely on scaffolds when they encounter particularly difficult problems. Providing scaffolds is a form of guided practice.

Scaffolds include modeling the steps by the teacher, or thinking aloud by the teacher as he or she solves the problem. Scaffolds also may be tools, such as cue cards or checklists, that complete part of the task for the students, or a model of the completed task against which students can compare their own work.

> One characteristic of effective teachers is their ability to anticipate students' errors and warn them about possible errors some of them are likely to make.

The process of helping students solve difficult problems by modeling and providing scaffolds has been called "cognitive apprenticeship." Students learn strategies and content during this apprenticeship that enable them to become competent readers, writers, and problem solvers. They are aided by a master who models, coaches, provides supports, and scaffolds them as they become independent.

In the classroom
One form of scaffolding is to give students prompts for steps they might use. Prompts such as "who," "why," and "how" have helped students learn to ask questions while they read. Teaching students to ask questions has been shown to help students' reading comprehension.

Similarly, one researcher developed the following prompt to help students organize material.[13]

1. Draw a central box and write the title of the article in it.
2. Skim the article to find four to six main ideas.
3. Write each main idea in a box below the central box.
4. Find and write two to four important details to list under each main idea.

Another form of scaffolding is thinking aloud by the teacher. For example, teachers might think aloud as they try to summarize a paragraph. They would show the thought processes they go through as they determine the topic of the paragraph and then use the topic to generate a summary sentence. Teachers might think aloud while solving a scientific equation or writing an essay, and at the same time provide labels for their mental processes. Such thinking aloud provides novice learners with a way to observe "expert thinking" that is usually hidden from the student. Teachers also can study their students' thought processes by asking them to think aloud during problem solving.

One characteristic of effective teachers is their ability to anticipate students' errors and warn them about possible errors some of them are likely to make. For example, a teacher might have students read a passage and then give them a poorly written topic sentence to correct. In teaching division or subtraction, the teacher may show and discuss with students the mistakes other students have frequently made.

In some of the studies, students were given a checklist to evaluate their work. Checklist items included "Have I found the most important information that tells me more about the main idea?" and "Does every sentence start with a capital letter?" The teacher then modeled use of the checklist.

In some studies, students were provided with expert models with which they could compare their work. For example, when students were taught to generate questions, they could compare their questions with those generated by the teacher. Similarly, when learning to write summaries, students could compare their summaries on a passage with those generated by an expert.

9. Require and monitor independent practice: Students need extensive, successful, independent practice in order for skills and knowledge to become automatic.

Research findings
In a typical teacher-led classroom, guided practice is followed by independent practice—by students working alone and practicing the new material. This independent practice is necessary because a good deal of practice (overlearning) is needed in order to become fluent and automatic in a skill. When material is overlearned, it can be recalled automatically and doesn't take up any space in working memory. When students become automatic in an area, they can then devote more of their attention to comprehension and application.

Independent practice provides students with the additional review and elaboration they need to become fluent. This need for fluency applies to facts, concepts, and discriminations that must be used in subsequent learning. Fluency is also needed in operations, such as dividing decimals, conjugating a regular verb in a foreign language, or completing and balancing a chemical equation.
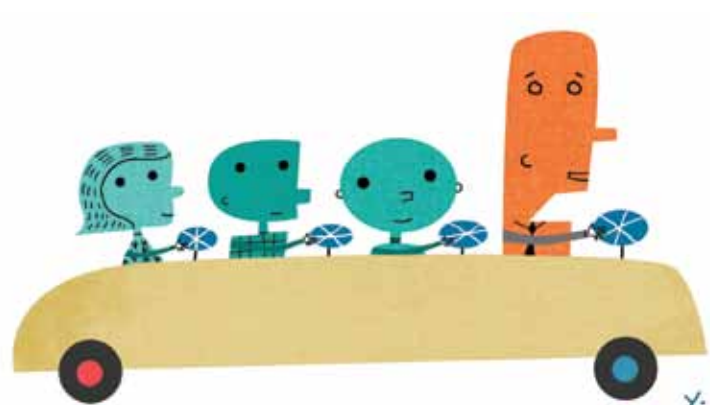
In the classroom

The more successful teachers provided for extensive and successful practice, both in the classroom and after class. Independent practice should involve the same material as the guided practice. If guided practice deals with identifying types of sentences, for example, then independent practice should deal with the same topic or, perhaps, with a slight variation, like creating individual compound and complex sentences. It would be inappropriate if the independent practice asked the students to do an activity such as "Write a paragraph using two compound and two complex sentences," however, because the students have not been adequately prepared for such an activity.

Students need to be fully prepared for their independent practice. Sometimes, it may be appropriate for a teacher to practice some of the seatwork problems with the entire class before students begin independent practice.

Research has found that students were more engaged when their teacher circulated around the room, and monitored and

> The best way to become an expert is through practice—thousands of hours of practice. The more the practice, the better the performance.

supervised their seatwork. The optimal time for these contacts was 30 seconds or less. Classrooms where the teachers had to stop at students' desks and provide a great deal of explanation during seatwork were the classrooms where students were making errors. These errors occurred because the guided practice was not sufficient for students to engage productively in independent practice. This reiterates the importance of adequately preparing students before they begin their independent practice.

Some investigators[14] have developed procedures, such as cooperative learning, during which students help each other as they study. Research has shown that all students tend to achieve more in these settings than do students in regular settings. Presumably, some of the advantage comes from having to explain the material to someone else and/or having someone else (other than the teacher) explain the material to the student. Cooperative learning offers an opportunity for students to get feedback from their peers about correct as well as incorrect responses, which promotes both engagement and learning. These cooperative/competitive settings are also valuable for helping slower students in a class by providing extra instruction for them.

10. Engage students in weekly and monthly review: Students need to be involved in extensive practice in order to develop well-connected and automatic knowledge.

Research findings

Students need extensive and broad reading, and extensive practice in order to develop well-connected networks of ideas (schemas) in their long-term memory. When one's knowledge on a particular topic is large and well connected, it is easier to learn new information and prior knowledge is more readily available for use. The more one rehearses and reviews information, the stronger these interconnections become. It is also easier to solve new problems when one has a rich, well-connected body of knowledge and strong ties among the connections. One of the goals of education is to help students develop extensive and available background knowledge.

Knowledge (even very extensive knowledge) stored in long-term memory that is organized into patterns only occupies a tiny amount of space in our limited working memory. So having larger and better-connected patterns of knowledge frees up space in our working memory. This available space can be used for reflecting on new information and for problem solving. The development of well-connected patterns (also called "unitization" and "chunking") and the freeing of space in the working memory is one of the hallmarks of an expert in a field.

Thus, research on cognitive processing supports the need for a teacher to assist students by providing for extensive reading of a variety of materials, frequent review, and discussion and application activities. The research on cognitive processing suggests that these classroom activities help students increase the number of pieces of information in their long-term memory and organize this information into patterns and chunks.

The more one rehearses and reviews information, the stronger the interconnections between the materials become. Review also helps students develop their new knowledge into patterns, and it

17 Principles of Effective Instruction

The following list of 17 principles emerges from the research discussed in the main article. It overlaps with, and offers slightly more detail than, the 10 principles used to organize that article.

- Begin a lesson with a short review of previous learning.
- Present new material in small steps with student practice after each step.
- Limit the amount of material students receive at one time.
- Give clear and detailed instructions and explanations.
- Ask a large number of questions and check for understanding.
- Provide a high level of active practice for all students.
- Guide students as they begin to practice.
- Think aloud and model steps.
- Provide models of worked-out problems.
- Ask students to explain what they have learned.
- Check the responses of all students.
- Provide systematic feedback and corrections.
- Use more time to provide explanations.
- Provide many examples.
- Reteach material when necessary.
- Prepare students for independent practice.
- Monitor students when they begin independent practice.

–B.R.

INTENTIONALLY
LEFT BLANK

Principles

(Continued from page 19)

helps them acquire the ability to recall past learning automatically.

The best way to become an expert is through practice—thousands of hours of practice. The more the practice, the better the performance.

In the classroom
Many successful programs, especially in the elementary grades, provided for extensive review. One way of achieving this goal is to review the previous week's work every Monday and the previous month's work every fourth Monday. Some effective teachers also gave tests after their reviews. Research has found that even at the secondary level, classes that had weekly quizzes scored better on final exams than did classes with only one or two quizzes during the term. These reviews and tests provided the additional practice students needed to become skilled, successful performers who could apply their knowledge and skills in new areas.

Teachers face a difficult problem when they need to cover a lot of material and don't feel they have the time for sufficient review. But the research states (and we all know from personal experience) that material that is not adequately practiced and reviewed is easily forgotten.

The 10 principles in this article come from three different sources: research on how the mind acquires and uses information, the instructional procedures that are used by the most successful teachers, and the procedures invented by researchers to help students learn difficult tasks. The research from each of these three sources has implications for classroom instruction, and these implications are described in each of these 10 principles.

Even though these principles come from three different sources, the instructional procedures that are taken from one source do not conflict with the instructional procedures that are taken from another source. Instead, the ideas from each of the sources overlap and add to each other. This overlap gives us faith that we are developing a valid and research-based understanding of the art of teaching. □

Endnotes

1. *Suggested readings*: George A. Miller, "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information," *Psychological Review* 63, no. 2 (1956): 81–97; and David LaBerge and S. Jay Samuels, "Toward a Theory of Automatic Information Processing in Reading," *Cognitive Psychology* 6, no. 2 (1974): 293–323.

2. *Suggested readings*: Carolyn M. Evertson, Charles W. Anderson, Linda M. Anderson, and Jere E. Brophy, "Relationships between Classroom Behaviors and Student Outcomes in Junior High Mathematics and English Classes," *American Educational Research Journal* 17, no. 1 (1980): 43–60; and Thomas L. Good and Jere E. Brophy, *Educational Psychology: A Realistic Approach*, 4th ed. (New York: Longman, 1990).

3. *Suggested readings*: Thomas L. Good and Douglas A. Grouws, "The Missouri Mathematics Effectiveness Project," *Journal of Educational Psychology* 71, no. 3 (1979): 355–362; and Alison King, "Guiding Knowledge Construction in the Classroom: Effects of Teaching Children How to Question and How to Explain," *American Educational Research Journal* 31, no. 2 (1994): 338–368.

4. *Suggested readings*: John Sweller, "Cognitive Load Theory, Learning Difficulty, and Instructional Design," *Learning and Instruction* 4, no. 4 (1994): 295–312; Barak Rosenshine, Carla Meister, and Saul Chapman, "Teaching Students to Generate Questions: A Review of the Intervention Studies," *Review of Educational Research* 66, no. 2 (1996): 181–221; and Alan H. Schoenfeld, *Mathematical Problem Solving* (New York: Academic Press, 1985).

5. *Suggested readings*: Evertson et al., "Relationships between Classroom Behaviors and Student Outcomes"; and Paul A. Kirschner, John Sweller, and Richard E. Clark, "Why Minimal Guidance during Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching," *Educational Psychologist* 41, no. 2 (2006): 75–86.

6. *Suggested readings*: Douglas Fisher and Nancy Frey, *Checking for Understanding: Formative Assessment Techniques for Your Classroom* (Alexandria, VA: Association for Supervision and Curriculum Development, 2007); and Michael J. Dunkin, "Student Characteristics, Classroom Processes, and Student Achievement," *Journal of Educational Psychology* 70, no. 6 (1978): 998–1009.

7. *Suggested readings*: Lorin W. Anderson and Robert B. Burns, "Values, Evidence, and Mastery Learning," *Review of Educational Research* 57, no. 2 (1987): 215–223; and Norman Frederiksen, "Implications of Cognitive Theory for Instruction in Problem Solving," *Review of Educational Research* 54, no. 3 (1984): 363–407.

8. *Suggested readings*: Michael Pressley and Vera Woloshyn, *Cognitive Strategy Instruction that Really Improves Children's Academic Performance*, 2nd ed. (Cambridge, MA: Brookline Books, 1995); and Barak Rosenshine and Carla Meister, "The Use of Scaffolds for Teaching Higher-Level Cognitive Strategies," *Educational Leadership* 49, no. 7 (April 1992): 26–33.

9. *Suggested readings*: Barak Rosenshine, "The Empirical Support for Direct Instruction," in *Constructivist Instruction: Success or Failure?* ed. Sigmund Tobias and Thomas M. Duffy (New York: Routledge, 2009), 201–220; and Robert E. Slavin, *Education for All* (Exton, PA: Swets and Zeitlinger, 1996).

10. *Suggested readings*: Good and Grouws, "The Missouri Mathematics Effectiveness Project"; and James A. Kulik and Chen-Lin C. Kulik, "College Teaching," in *Research on Teaching: Concepts, Findings, and Implications*, ed. Penelope L. Peterson and Herbert J. Walberg (Berkeley, CA: McCutchan, 1979).

11. Good and Grouws, "The Missouri Mathematics Effectiveness Project."

12. These stems were developed by King, "Guiding Knowledge Construction in the Classroom."

13. Sandra J. Berkowitz, "Effects of Instruction in Text Organization on Sixth-Grade Students' Memory for Expository Reading," *Reading Research Quarterly* 21, no. 2 (1986): 161–178. For additional strategies to help students organize material, see Wisconsin Department of Public Instruction, *Strategic Learning in the Content Areas* (Madison, WI: Wisconsin Department of Public Instruction, 2005).

14. Slavin, *Education for All*.

Approaches to Evaluating
Teacher Effectiveness:
A Research Synthesis

June 2008

Laura Goe, Ph.D.
Courtney Bell, Ph.D.
Olivia Little
ETS

3018_08/08

Acknowledgments

Contents

Abstract

This research synthesis examines how teacher effectiveness is currently measured. By evaluating the research on teacher effectiveness and the different instruments used to measure it, this research synthesis contributes to the discussion of appropriate rigor and relevance of measures for different purposes (i.e., formative vs. summative evaluation). The findings are presented along with related policy implications. In addition, the synthesis describes how various measures have been evaluated, explains why certain measures are most suitable for certain purposes (high-stakes evaluation vs. formative evaluation, for instance), and suggests how the results of the study might be used to inform the national conversation about teacher effectiveness. A comprehensive definition of the components and indicators that characterize effective teachers is provided, extending this definition beyond teachers' contribution to student achievement gains to include how teachers impact classrooms, schools, and their colleagues as well as how they contribute to other important outcomes for students. Through this synthesis, the National Comprehensive Center for Teacher Quality (TQ Center) hopes to provide some practical guidance in how best to evaluate teacher effectiveness.

Introduction

The process of evaluating the effectiveness of teachers has changed over time along with the definition of what effective teaching is, due in part to increasing state and federal attention to school-level and classroom-level accountability for student learning. Effective teaching has been defined in many ways throughout the years (Campbell, Kyriakides, Muijs, & Robinson, 2003; Cheng & Tsui, 1999; Cruickshank & Haefele, 1990; Good, 1996; Muijs, 2006), and methods for measuring teachers have changed as definitions and beliefs about what is important to measure have evolved. Although there is a general consensus that good teaching matters and that it may be the single most important school-based factor in improving student achievement (Darling-Hammond, 2000; Wright, Horn, & Sanders, 1997), measuring teacher effectiveness has remained elusive in part because of ongoing debate about what an effective teacher is and does. In a discussion of research-based indicators of effective teaching, Cruickshank and Haefele (1990) stated, "An enormous underlying problem with teacher evaluation relates to lack of agreement about what constitutes good or effective teaching" (p. 34).

Besides a lack of clear consensus on what an effective teacher is and does—or perhaps because of it—there is not a generally agreed-upon method for evaluating teacher effectiveness. Commonly used methods include classroom observations designed to measure teacher practices against some standard of effective teaching and value-added models that set out to measure the contribution of individual teachers to their students' achievement gains. This research synthesis, describes the various ways in which effective teaching can be conceptualized and measured and consists of the following sections:

- Rationale and Goals of This Study
- Important Definitions and Specifications
- Proposal of a Comprehensive Definition of Teacher Effectiveness
- Data Collection and Methods
- Validity and Considerations in Measuring Teacher Effectiveness
- Methods of Measuring Teacher Effectiveness
- Considering a Comprehensive Measure of Teacher Effectiveness
- Policy Recommendations and Implications

Rationale and Goals of This Study

The primary goal of this research synthesis is to help regional and state decision makers better understand what constitutes effective teaching and the advantages and disadvantages of the various measures commonly used to evaluate it. This study was commissioned by the National Comprehensive Center for Teacher Quality (TQ Center), which is charged with assisting regional comprehensive centers and the states they work with to understand and implement the highly qualified teacher requirements of the No Child Left Behind (NCLB) Act, with a particular emphasis on ensuring that students at risk for poor educational outcomes and students with special needs are served by highly qualified, effective teachers.

The TQ Center gathers data regularly to determine the most pressing needs of the states in regard to implementing the NCLB highly qualified teacher requirements. The TQ Center's needs-sensing data strongly suggest that states need more help identifying effective teachers in order to better respond to the NCLB equitable distribution requirement, which states that minority students and students living in poverty must have equal access to experienced, highly qualified teachers. This requirement focuses on ensuring equal access to highly qualified, experienced teachers; however, all students, particularly those who are at high risk for failure, also should have access to effective teachers. Mandating that teachers meet the minimum requirements to be considered highly qualified is a first step toward ensuring teacher effectiveness, but just meeting those requirements is no guarantee that teachers will be effective (Goe, 2007; Gordon, Kane, & Staiger, 2006).

The topic of this research synthesis is central to the mission of the TQ Center. A research-based discussion of teacher effectiveness—its definition and measurement as well as the advantages and drawbacks of different ways of measuring teacher effectiveness—can help inform states as they develop their own mechanisms for establishing teacher effectiveness more directly.

Important Definitions and Specifications

Evaluating teachers can be approached from three different but related angles: measurement of inputs, processes, and outputs. *Inputs* are what a teacher brings to his or her position, generally measured as teacher background, beliefs, expectations, experience, pedagogical and content knowledge, certification and licensure, and educational attainment. These measures are sometimes discussed in the literature as "teacher quality"; for instance, the NCLB requirement for highly qualified teachers refers specifically to teacher qualifications and credentials. *Processes,* on the other hand, refers to the interaction that occurs in a classroom between teachers and students. It also may include a teacher's professional activities within the larger school and community, but for the purposes of this research synthesis, classroom processes are the focus. *Outputs* represent the results of classroom processes, such as impact on student achievement, graduation rates, student behavior, engagement, attitudes, and social-emotional well-being. Other outcomes may involve contributions to the school or community in the form of taking on school leadership roles, educating other teachers, or strengthening relationships with parents, but again for the purposes of this research synthesis, student outcomes are the focus. Outputs can be referred to as "teacher effectiveness," although as discussed in the following section, *teacher effectiveness* as used in the research literature is often limited to mean impact on student achievement specifically.

The studies discussed in this research synthesis focus explicitly on teacher effectiveness in terms of gains in student achievement and on measures of classroom processes. The reasons for using this focus and selection criteria are described in the Data and Methods section. However, given the many terms discussed and their subtle distinctions, an argument can be made for a conceptualization of *teacher effectiveness* that is a broader and more encompassing term for the many facets that contribute to a teacher's success.

Defining Teacher Effectiveness

Clarifying the way *teacher effectiveness* is defined is important for two main reasons. First, what is measured is a reflection of what is valued, and as a corollary, what is measured is valued. Definitions nominate and shape what needs to be measured. If, for example, policy conversations revolve around scores from standardized tests, the significant outcomes can be narrowed to those that can be measured with standardized test scores. On the other hand, when policy conversations concern the interactions between teachers and students, the focus shifts to classrooms and documenting effective interactions among teachers and their students. In addition, different definitions lead to different policy solutions. When the conversation focuses on teacher quality, the discussion likely turns to improving teachers' scores on measures of knowledge or on signals of that knowledge, such as certification. When classroom processes are discussed, particular practices or approaches to teaching become the focus.

Given the importance of these distinctions, this research synthesis uses the term *teacher effectiveness* but does so with a much broader definition than is typically associated with that term in current policy conversations. In the remainder of this section, a more nuanced definition of *teacher effectiveness* is provided; this definition includes the varied roles teachers play as well as the varied student outcomes education stakeholders value.

Critiques of the Dominant Teacher Effectiveness Definition

Increasingly, policy conversations frame *teacher effectiveness* as a teacher's ability to produce higher than expected gains in students' standardized test scores. This focus on attributing gains on standardized tests to teachers and measuring the result of teaching by averaging test score gains has a number of strengths. It is parsimonious; it can be measured using data collected as part of NCLB requirements; and it has a certain amount of credibility—most would agree that an effective teacher *should* help students learn more than expected. This definition does, however, have serious limitations.

Teachers Are Not Solely Responsible for Students' Learning.

One critique concerns the problem of the assumptions of causality that underlie this approach. The approach requires the establishment of what part of an effectiveness score is attributable solely to the teacher. Making this determination is problematic not just for practical reasons but for logical reasons—assumptions are required that may be unreasonable. Fenstermacher and Richardson (2005) illustrate the problem with this scenario:

> If we presuppose a blank, receptive mind, encased within a compliant and passive learner, then we need travel only a very short logical distance to infer that teaching produces learning, and hence that what teachers do determines whether students learn. In the passive recipient view, it makes some sense to think of successful teaching arising solely from the actions of a teacher. That is, learning on the part of the student is indeed a direct result of actions by a teacher. Yet we all know that learners are not passive receptors of information directed at them. Learning does not arise solely on the basis of teacher activity. Assuming that the formulation offered above has merit, then it follows that success at learning requires a combination of circumstances well beyond the actions of a teacher. (pp. 190–191)

It can be argued that narrowing the definition of teacher effectiveness to reflect only student growth on standardized achievement measures takes this assumption too far. It is important to note that measures of teacher effectiveness can be calculated without regard to what takes place in classrooms and schools, if teacher effectiveness is narrowly defined as a given teacher's impact on the learning of his or her students as measured by standardized tests. With this narrow definition, other important ways that teachers contribute to successful students, communities, and schools are overlooked. Similarly, other influences on student outcomes, including other teachers, peers, school resources, community support, leadership, and school climate or culture, cannot be "parceled out" of the resulting score.

In the narrowest definition of teacher effectiveness, in which effectiveness is determined solely by student achievement gains, a teacher can be deemed effective compared to other teachers because his or her students performed better on the state test than the students' prior achievement would have predicted, without consideration of any other factors. In that case, it would be impossible to say whether the growth in achievement as reflected by test scores was the result of class time spent narrowly on test-taking skills and test preparation activities or whether

achievement growth was the result of inspired, competent teaching of a broad, rich curriculum that engaged students, motivated their learning, and prepared them for continued success.

Consensus Should Drive Research, Not Measurement Innovations.

Another critique of a teacher effectiveness model based on test scores concerns the degree to which innovations in measurement drive how teacher effectiveness is defined. Campbell et al. (2003) contend that trends in measurement of teacher effectiveness seem to follow the development of new instruments and technologies, focusing on the ability to measure something, rather than first defining effectiveness and *then* determining a technology for measuring it. They describe the sense of "…the horse and the cart being in the wrong places; the technology of measurement has been creating the concept of effectiveness rather than the concept requiring an appropriate technology. It follows that current concepts of teacher effectiveness may be open to question" (p. 350). These authors make an important point: just because it is possible to match teachers to their students' test scores and use this relationship as a measure of teacher effectiveness does not mean that this is the *only* way to evaluate teacher effectiveness.

The increased availability of data in which student achievement is linked to teachers along with statistical innovations in analyzing these data may be partly responsible for what appears to be a growing emphasis on measuring teachers' contributions to student achievement (Drury & Doran, 2003; Hershberg, Simon, & Lea-Kruger, 2004; The Teaching Commission, 2004) and a concomitant narrowing of the definition of teacher effectiveness. Students' knowledge is summarized in a test score, whereas teachers' effectiveness is reflected in their contribution to that test score.

Value-added models provide a classic example of a measure of teacher effectiveness driven by technological development. Using longitudinal linked teacher-student data, William Sanders was able to determine that students in some teachers' classrooms were scoring higher than their previous test scores would have predicted (Sanders & Rivers, 1996). Sanders' findings and his marketing of the technology to states for the purpose of evaluating schools and teachers have garnered considerable attention and contributed to the increased use of value-added methodologies.

In addition to the objection to innovations leading definitions, there are substantial issues with using student achievement test scores as measures of teaching effectiveness for all students. If, for example, students are dropping out of school at a higher rate because of testing-related graduation requirements, as some research suggests (Haney, 2000), then high school achievement scores are increasingly representing the scores of the "survivors" rather than *all* students. Such measurement issues raise questions about the validity of test scores as a measure of teacher effectiveness in secondary schools with high dropout rates.

Learning Is More Than Average Achievement Gains.

A final critique of this model suggests that an overly narrow focus on standardized test scores as the most important—and in some cases, only—student outcome measure is not aligned with what the field agrees an effective teacher does. Though current policy conversations and some

research studies implicitly refer to teacher effectiveness as gains in student achievement, reviewing the literature on teacher evaluation revealed that definitions of teacher effectiveness provided by researchers have been more varied and broader in scope. For example, Campbell, Kyriakides, Muijs, and Robinson (2004) state, "Teacher effectiveness is the impact that classroom factors, such as teaching methods, teacher expectations, classroom organisation, and use of classroom resources, have on students' performance" (p. 3). This definition takes into consideration what occurs in the classroom, but the measure of effectiveness is still the students' performance. However, a number of researchers contend that there are other important outcomes besides students' performance on standardized tests that define effective teachers. More than 20 years ago, in their review of "process-outcome" research linking teacher behavior to student achievement, Brophy and Good (1986) made the following statement about their work:

> The research discussed is concerned with teachers' effects on students, but it is a misnomer to refer to it as "teacher effectiveness" research, because this equates "effectiveness" with success in producing achievement gain. What constitutes "teacher effectiveness" is a matter of definition, and most definitions include success in socializing students and promoting their affective and personal development in addition to success in fostering their mastery of formal curricula. (p. 328)

Brophy and Good's point becomes clear when the outcome measure of graduation is considered. In *A Highly Qualified Teacher in Every Classroom: The Secretary's Fourth Annual Report on Teacher Quality*, it is clear that improving graduation rates is an important goal that is tied to teaching: "While much of the work of NCLB has focused on elementary and middle schools, now, America must do more to prepare high school students for graduation, especially those most at risk of dropping out" (Office of Postsecondary Education, 2005, p. xii). Yet even though on-time promotion and high school graduation are important educational outcomes, they are ignored under an achievement-only definition of teacher effectiveness.

It could be that standards for judging effectiveness have become more focused on student achievement as the most important outcome due to increasing accountability pressures. Or it could be that the accessibility of linked student-teacher data, improvements in statistical methods, and increasingly powerful computers have made it possible to do analyses that were previously extremely difficult to perform. Most likely, it is a combination of those factors. Student achievement gains should be an important component in evaluating teacher effectiveness; however, the critiques of the achievement-focused view of teacher effectiveness are legitimate. The next section offers a broader view of teacher effectiveness and argues that other aspects of teaching must be a part of the conversation.

Considering a More Comprehensive Definition
of Teacher Effectiveness

In light of these critiques, and given that teachers' roles involve much more than simply providing subject-matter instruction, it is appropriate to consider a broader and more comprehensive definition of effective teachers consisting of five points and formulated by evaluating discussions of teacher effectiveness in the research literature as well as in policy documents, standards, and reports (e.g., Berry, 2004; Brophy & Good, 1986; Campbell et al., 2003, 2004; Cheng & Tsui, 1999; Darling-Hammond & Youngs, 2002; Englert, Tarrant, & Mariage, 1992; Fenstermacher & Richardson, 2005; Gentilucci, 2004; Hamre & Pianta, 2005; Haycock, 2004; Interstate New Teacher Assessment and Support Consortium, 2001; Kyriakides, 2005; McCaffrey, Lockwood, Koretz, & Hamilton, 2003; McColskey et al., 2005; Muijs, 2006; National Board for Professional Teaching Standards, 2002; Newmann, Bryk, & Nagaoka, 2001; Odden, Borman, & Fermanich, 2004; Office of Postsecondary Education & Office of Policy Planning and Innovation, 2003; Rivkin, Hanushek, & Kain, 2005; Schlusmans, 1978; Shavelson, Webb, & Burstein, 1986; Tucker & Stronge, 2005; Vandevoort, Amrein-Beardsley, & Berliner, 2004; Watson & De Geest, 2005). In addition, after these five points were conceptualized, they were circulated among a number of experts on teacher quality and effectiveness for feedback and strengthened as a result the experts' input.

The five-point definition of effective teachers consists of the following:

- Effective teachers have high expectations for all students and help students learn, as measured by value-added or other test-based growth measures, or by alternative measures.

- Effective teachers contribute to positive academic, attitudinal, and social outcomes for students such as regular attendance, on-time promotion to the next grade, on-time graduation, self-efficacy, and cooperative behavior.

- Effective teachers use diverse resources to plan and structure engaging learning opportunities; monitor student progress formatively, adapting instruction as needed; and evaluate learning using multiple sources of evidence.

- Effective teachers contribute to the development of classrooms and schools that value diversity and civic-mindedness.

- Effective teachers collaborate with other teachers, administrators, parents, and education professionals to ensure student success, particularly the success of students with special needs and those at high risk for failure.

This definition is intended to focus measurement efforts on multiple components of teacher effectiveness. It is proposed not as a criticism of other useful definitions, many of which were considered in the formation of these points, but as a means of *clarifying priorities* for measuring teaching effectiveness. The first point directly addresses student achievement gains on standardized tests, and the other points focus on teachers' contributions that may ultimately improve student learning, albeit indirectly. Clearly, student achievement gains on standardized tests are not the only—possibly not even the most important—outcome against which teacher

performance should be evaluated. A comprehensive evaluation of teacher effectiveness might be based on a composite that includes teachers' scores using a number of different measures.

Some may argue that teacher effectiveness should be limited to outcome measures, and thus process and behavior variables (e.g., having high expectations, using appropriate assessments, or collaborating with parents) should be excluded. However, because teachers impact student learning and growth through the processes and practices they employ, it is reasonable to state that an effective teacher can be observed to be doing things that research has suggested are likely to lead to improved student learning. It is necessary for these processes and practices to be measurable.

Although it is theoretically possible to identify indicators of all the components in the definition of effective teachers so that they can be measured and scored, there is a dearth of research in many of these areas. Most measures of teacher effectiveness focus on either student achievement gains attributed to the teacher *or* on classroom performance as measured with observation protocols. Actually *measuring* teachers' contribution to other outcomes—student attendance, promotion, and graduation—is less common. The fifth point in the definition is seldom measured or even considered as a component of teacher effectiveness, but it is particularly important given the increased emphasis on collaboration between general education teachers and those who focus on working with students with special needs (e.g., Abbott, Walton, Tapia, & Greenwood, 1999; Bauer, Johnson, & Sapona, 2004; Benner & Judge, 2000; Blanton, Blanton, & Cross, 1994; Blanton, Griffin, Winn, & Pugach, 1997; Fuchs & Fuchs, 1998; Gable, 1993; Hardman, McDonnell, & Welch, 1998; Interstate New Teacher Assessment and Support Consortium, 2001; Pugach, 2005). The next section describes the process through which the literature was selected and narrowed down in order to present information about various ways that teaching is measured and to make suggestions about how teacher effectiveness can be more comprehensively measured.

Data Collection and Methods

General Approach

The general approach to the identification and selection of articles for this synthesis was to start with broad categories and many search terms and then progressively narrow the group of studies down to only those that met certain criteria. While stricter criteria could have been applied, the authors of this synthesis are in agreement with Dynarksi (2008) who states, "Selective exclusion of research requires great caution, as selectivity can be interpreted as compromising scientific objectivity for purposes that educators cannot discern and may misinterpret" (p. 27). Rather than eliminate studies that might be informative for some purposes or audiences, the authors of this synthesis elected not to use narrow criteria. Dynarski also stated:

> Certainly it is possible that the findings from some studies are due to publication bias or arise from local conditions that are unusual or hard to replicate. But if syntheses review all the evidence and apply sound standards, educators can make up their own minds about whether the findings are credible or whether the implementation conditions are unrealistic and not useful to them. (p. 28)

Given that the purpose of this synthesis is to help policymakers, state leaders, and educational professionals sort out what the evidence says about teacher effectiveness, it seemed reasonable to let them weigh the evidence for themselves.

Stages of Development

Several stages were required to develop an appropriate set of articles to analyze for this synthesis. The authors served as reviewers of all articles and made decisions at each stage of the process based on their shared understanding of the identified criteria. In the case that one author was uncertain about whether an article met the criteria, she consulted with one of the other authors and discussed the uncertainty until a consensus was reached.

It is worth noting, however, that the literature on teacher effectiveness is large and disconnected. Scholars working in different fields theorize, conduct studies, and publish articles in very different journals. Sometimes these findings do not build on or connect with findings in other areas. This can mean that knowledge is less cumulative than one might like. As Kennedy (2007) notes, this means that reviews of research in such areas rely on the conceptual frameworks of the researchers. The authors of this research synthesis selected categories that they deemed to be reasonable; however, scholars in other disciplines might have used different categories.

Stage 1

The authors met on a number of occasions to discuss the purpose of the synthesis and develop a list of search terms that appeared to fit with that purpose.

Stage 2

Articles were identified through Internet and library searches of keywords and phrases related to the topics of teacher effectiveness and measuring teacher performance. ERIC and PsycInfo were the main databases used to identify relevant peer-reviewed articles within the last six to eight years, using the following search terms: *teacher effectiveness*, *teacher evaluation*, *value-added modeling*, *teaching methods*, *teacher improvement*, *teacher competencies*, *pedagogical content knowledge*, *instructional effectiveness*, *instructional improvement*, *research tools*, *videotape recordings*, *questionnaires*, *instructional material evaluation*, *teacher behavior*, *assignments*, *instructional development*, *beginning teacher induction*, *professional development*, *academic achievement prediction*, *educational measurement*, and *educational quality*. Additional articles, including older, seminal, nonempirical, and/or theoretical pieces, were identified from broader Internet searches, reference lists of related articles, and recommendations of experts in the field.

Stage 3

This search process yielded more than 1,600 studies. In order to narrow the results further, abstracts were reviewed to determine whether the studies met the following criteria:

- **Language and Location.** Studies were published in English, and research was conducted in the United States, Canada, Great Britain, Ireland, Australia, and New Zealand.

- **Population.** Research addressed the K–12 student population and measured inservice teachers.

- **Relevance.** Research addressed the topic of measuring effective teaching.

Approximately 300 articles meeting these criteria were then sent to the next stage.

Stage 4

The remaining 300 articles were reviewed more closely for relevance and methodological rigor. Studies chosen for this research synthesis met the following additional criteria:

- They were empirical.
- They included a measure of teacher effectiveness or classroom practice.
- They included a student outcome measure *or* had implications for teacher effectiveness.
- They reported methods meeting accepted standards for quality research (e.g., reliable and validated instruments, appropriate study design, and necessary controls).

Stage 5

The resulting collection of studies was then evaluated, and additional exclusions were made when deeper reading of studies revealed they did not meet the purposes or the quality standards of this synthesis. Studies that were of poor quality, off topic, out-of-scope, focused on higher education or prekindergarten education, or lacked descriptions of data and methods were

excluded. The resulting synthesis includes approximately 120 studies that were thoroughly reviewed.

As discussed, the search was narrowed by focusing on studies measuring classroom processes and outputs in the form of student outcomes, paying particular attention to studies measuring teacher effectiveness in terms of value-added student achievement measures. The search was limited in this way for two main reasons:

- A previous research synthesis commissioned by the TQ Center (see Goe, 2007) specifically addresses the links between measures of teacher quality and student outcomes, and this topic also has been addressed in a number of other research syntheses and reviews (e.g., Darling-Hammond & Youngs, 2002; Goe, 2007; Rice, 2003; Wayne & Youngs, 2003; Wilson & Floden, 2003). Though there is some overlap, this research synthesis is meant to be an extension of previous work, thus it focuses on processes and outputs rather than on inputs.

- The criteria was narrowed by only including processes occurring inside the classroom and outputs concerning student outcomes. This narrowing of scope was necessary to ensure that the amount of literature to be reviewed and synthesized was manageable enough to be transformed into a useable and informative document. The research synthesis mainly focuses on processes inside the classroom and student outcomes related to gains in student achievement because these are topics that are prevalent in the current education policy landscape and are areas in which states have indicated a need for more information and assistance.

Furthermore, this synthesis is limited to measuring teachers and does not address methods of measuring school effects, the effectiveness of curriculum or professional development implementations (unless they include measures specific to teachers), or other evaluations of educational interventions or programming. Though these are important and related topics, they are beyond the scope of this synthesis.

Validity and Considerations in Measuring Teacher Effectiveness

Determining what type of teacher evaluation method is best for a given purpose includes taking account of the validity and reliability of the instrument or process being used. Validity is the "most fundamental consideration in assuring the quality of any assessment" (Millett, Stickler, Payne, & Dwyer, 2007, p. 4). Validity refers to the degree to which an interpretation of a test score, or in this case, a score from a measure of teacher effectiveness, is supported by evidence. For a measure of teacher effectiveness to be valid, evidence must support the argument that the measure actually assesses the dimension of teacher effectiveness it claims to measure and not something else. In addition, evidence that the measure is valid for the purpose for which it will be used is essential. Instruments cannot be valid in and of themselves; an instrument or assessment must be validated for particular purposes (Kane, 2006; Messick, 1989). For example, an observation-based score might be validated for professional development purposes but might not be validated for compensation purposes. Determining the validity of an instrument requires taking account of the evidence regarding what the instrument measures, what it does not measure, and how the scores are being used. This requires the user of the instrument to be well-informed about these issues and willing to make judgments about the degree to which there is sufficient evidence to use a particular instrument for the purpose under consideration.

In addition to concerns about validity, there are other measurement concerns. Blanton et al. (2003) identified six criteria that are particularly useful in informing this conversation [which are elaborated in Coggshall (2007)], and these criteria have been adapted and applied to the discussion of teacher effectiveness in the following pages.

Comprehensiveness refers to the degree to which a measure captures *all* of the various aspects of teacher effectiveness. For example, less comprehensive measures might only capture how well a teacher is able to represent mathematics in the classroom. More comprehensive measures would capture how teachers represent mathematics, how they scaffold student learning, and how well they work with colleagues.

Generality refers to how well an instrument captures the full range of contexts in which teachers work. If an instrument can be used to assess elementary and secondary teacher effectiveness in music and special education, the instrument can be said to have a high level of generality. Generality is particularly important if one intends to compare teachers across contexts.

Utility refers to how useful scores from an instrument are for a specific purpose. For example, scores from an instrument that ignores teaching context may not be useful in identifying contexts that appear to support more effective teaching. The experience of other researchers or practitioners with an instrument makes it possible to better anticipate its potential uses and limitations.

Practicality refers to the logistical issues associated with a measure. These include the "costs, training requirements, and the developmental work required to adapt an existing model or measure" for one's own purpose (Blanton et al., 2003, p. 14). For example, creating valid and reliable instruments and processes for measuring teacher effectiveness is costly and time-

consuming. Adapting an existing instrument and process might be less of a drain on district or state resources.

Reliability refers to the degree to which an instrument measures something consistently. For example, it might be important to know whether scores on an instrument measuring teacher effectiveness vary by time of year, time of day, grade level, or subject matter. It is also important to note that instruments can be reliable without actually measuring what they were intended to measure. For example, an instrument might consistently measure teachers' use of flash cards. But if flash card use is not an important determinant of teacher effectiveness, then the instrument is reliable but not valid for the purpose of measuring teacher effectiveness.

Credibility is a specific type of validity—face validity—that is particularly important in measures of teacher effectiveness. If an instrument has strong credibility, many stakeholders from different groups (e.g., parents, teachers, administrators, and policymakers) view the measure as reasonable and appropriate.

In this research synthesis, these aspects of measurement—validity, comprehensiveness, generality, utility, practicality, reliability, and credibility—are used to describe and assess a range of approaches to measuring teacher effectiveness. Particular attention is given to issues of validity and reliability because the authors draw heavily from the research literature, which is very concerned with such issues.

In addition, careful attention is given to the purposes of instruments. The authors distinguish between high-stakes, low-stakes, formative, and summative assessments of teacher effectiveness. A formative evaluation is one that is intended to gather information that will be useful to improve a program, activity, or behavior. A summative evaluation is meant to make a final determination about a program, activity, or behavior at a specific point in time. For instance, a classroom observation may be an informal drop-in visit by a principal, or it may be a planned, formal observation conducted by highly trained professional evaluators with employment or tenure consequences. An informal evaluation that does not carry serious consequences and is meant to collect information for providing feedback to improve teaching is considered low-stakes and formative. In contrast, formal evaluations that carry substantial consequences and are conducted to gather information for a specific decision-making process are considered high-stakes and summative. Considering whether the intent of the evaluation is high-stakes or low-stakes and whether it is summative or formative in nature will have strong implications for choosing a measure that will provide valid results.

Methods of Measuring Teacher Effectiveness

The following sections present methods in teacher evaluation that are useful for measuring teacher effectiveness more broadly and providing information about what makes teachers effective. The discussion begins with the most widely used measure of teacher effectiveness, classroom observations. A review of other instruments that directly assess what teachers do in classrooms also is provided. These include principal evaluations; analysis of classroom artifacts (i.e., ratings of teacher assignments and student work); teaching portfolios; teacher self-reports of practice, including surveys, teaching logs, and interviews; and student ratings of teacher performance. Finally, teacher effectiveness as measured by value-added strategies is considered. For the scope of this discussion, more indirect measures of teaching, such as teacher demonstrations of knowledge, teacher responses to theoretical teaching situations (i.e., structured vignettes), or parent satisfaction surveys are not included. These measures can be extremely useful in assessing teaching competency; however, the authors chose to focus on measures that more directly assess the processes and activities occurring during instruction and products that are created inside the classroom. In addition, the research linking credentials, experience, or knowledge to teacher effectiveness is not considered. Though such work is terrifically important in discussions of initial teacher licensure, extensive reviews have already been conducted and widely publicized (e.g., Darling-Hammond & Youngs, 2002; Goe, 2007; Rice, 2003; Wayne & Youngs, 2003; Wilson & Floden, 2003).

Each of the sections that follow defines and describes the measure, provides examples and research findings on its use, and discusses its strengths and cautions, keeping in mind the previously described validity considerations and providing recommendations as appropriate. Coverage of instruments is not meant to be exhaustive but rather to accomplish the following: (1) to provide some researched examples of methods that are being employed by states or that are promising measures of teaching, and (2) to present knowledge of their uses and barriers. In addition, many commercially available products are not reviewed here but are examples of the broader class of instruments considered in this synthesis. Thus, in the interest of time, the synthesis considers the broader class of instruments and leaves it to the reader to consider the particular products. Table 1 presents a brief summary of the discussion on each method.

Table 1. Brief Summaries of Teacher Evaluation Methods

Measure	Description	Research	Strengths	Cautions
Classroom Observation	Used to measure observable classroom processes, including specific teacher practices, holistic aspects of instruction, and interactions between teachers and students. Can measure broad, overarching aspects of teaching or subject-specific or context-specific aspects of practice.	Some highly researched protocols have been found to link to student achievement, though associations are sometimes modest. Research and validity findings are highly dependent on the instrument used, sampling procedures, and training of raters. There is a lack of research on observation protocols as used in context for teacher evaluation.	• Provides rich information about classroom behaviors and activities. • Is generally considered a fair and direct measure by stakeholders. • Depending on the protocol, can be used in various subjects, grades, and contexts. • Can provide information useful for both formative and summative purposes.	• Careful attention must be paid to choosing or creating a valid and reliable protocol and training and calibrating raters. • Classroom observation is expensive due to cost of observers' time; intensive training and calibrating of observers adds to expense but is necessary for validity. • This method assesses observable classroom behaviors but is not as useful for assessing beliefs, feelings, intentions, or out-of-classroom activities.
Principal Evaluation	Is generally based on classroom observation, may be structured or unstructured; uses and procedures vary widely by district. Is generally used for summative purposes, most commonly for tenure or dismissal decisions for beginning teachers.	Studies comparing subjective principal ratings to student achievement find mixed results. Little evidence exists on validity of evaluations as they occur in schools, but evidence exists that training for principals is limited and rare, which would impair validity of their evaluations.	• Can represent a useful perspective based on principals' knowledge of school and context. • Is generally feasible and can be one useful component in a system used to make summative judgments and provide formative feedback.	• Evaluation instruments used without proper training or regard for their intended purpose will impair validity. • Principals may not be qualified to evaluate teachers on measures highly specialized for certain subjects or contexts.

Measure	Description	Research	Strengths	Cautions
Instructional Artifact	Structured protocols used to analyze classroom artifacts in order to determine the quality of instruction in a classroom. May include lesson plans, teacher assignments, assessments, scoring rubrics, and student work.	Pilot research has linked artifact ratings to observed measures of practice, quality of student work, and student achievement gains. More work is needed to establish scoring reliability and determine the ideal amount of work to sample. Lack of research exists on use of structured artifact analysis in practice.	• Can be a useful measure of instructional quality if a validated protocol is used, if raters are well-trained for reliability, and if assignments show sufficient variation in quality. • Is practical and feasible because artifacts have already been created for the classroom.	• More validity and reliability research is needed. • Training knowledgeable scorers can be costly but is necessary to ensure validity. • This method may be a promising middle ground in terms of feasibility and validity between full observation and less direct measures such as self-report.
Portfolio	Used to document a large range of teaching behaviors and responsibilities. Has been used widely in teacher education programs and in states for assessing the performance of teacher candidates and beginning teachers.	Research on validity and reliability is ongoing, and concerns have been raised about consistency/stability in scoring. There is a lack of research linking portfolios to student achievement. Some studies have linked NBPTS certification (which includes a portfolio) to student achievement, but other studies have found no relationship.	• Is comprehensive and can measure aspects of teaching that are not readily observable in the classroom. • Can be used with teachers of all fields. • Provides a high level of credibility among stakeholders. • Is a good tool for teacher reflection and improvement.	• This method is time-consuming on the part of teachers and scorers; scorers should have content knowledge of the portfolios. • The stability of scores may not be high enough to use for high-stakes assessment. • Portfolios are difficult to standardize (compare across teachers or schools). • Portfolios represent teachers' exemplary work but may not reflect everyday classroom activities.

Measure	Description	Research	Strengths	Cautions
Teacher Self-Report Measure	Teacher reports of what they are doing in classrooms. May be assessed through surveys, instructional logs, and interviews. Can vary widely in focus and level of detail.	Studies on the validity of teacher self-report measures present mixed results. Highly detailed measures of practice may be better able to capture actual teaching practices but may be harder to establish reliability or may result in very narrowly focused measures.	• Can measure unobservable factors that may affect teaching, such as knowledge, intentions, expectations, and beliefs. • Provides the unique perspective of the teacher. • Is very feasible and cost-efficient; can collect large amounts of information at once.	• Reliability and validity of self-report is not fully established and depends on instrument used. • Using or creating a well-developed and validated instrument will decrease cost-efficiency but will increase accuracy of findings. • This method should not be used as a sole or primary measure in teacher evaluation.
Student Survey	Used to gather student opinions or judgments about teaching practice as part of teacher evaluation and to provide information about teaching as it is perceived by students.	Several studies have shown that student ratings of teachers can be useful in providing information about teaching; may be as valid as judgments made by college students and other groups; and, in some cases, may correlate with measures of student achievement. Validity is dependent on the instrument used and its administration and is generally recommended for formative use only.	• Provides perspective of students who have the most experience with teachers. • Can provide formative information to help teachers improve practice in a way that will connect with students. • Makes use of students, who may be as capable as adult raters at providing accurate ratings.	• Student ratings have not been validated for use in summative assessment and should not be used as a sole or primary measure of teacher evaluation. • Students cannot provide information on aspects of teaching such as a teacher's content knowledge, curriculum fulfillment, and professional activities.

Measure	Description	Research	Strengths	Cautions
Value-Added Model	Used to determine teachers' contributions to students' test score gains. May also be used as a research tool (e.g., determining the distribution of "effective" teachers by student or school characteristics).	Little is known about the validity of value-added scores for identifying effective *teaching*, though research using value-added models does suggest that teachers differ markedly in their contributions to students' test score gains. However, correlating value-added scores with teacher qualifications, characteristics, or practices has yielded mixed results and few significant findings. Thus, it is obvious that teachers vary in effectiveness, but the reasons for this are not known.	• Provides a way to evaluate teachers' contribution to student learning, which most measures do not. • Requires no classroom visits because linked student/teacher data can be analyzed at a distance. • Entails little burden at the classroom or school level because most data is already collected for NCLB purposes. • May be useful for identifying outstanding teachers whose classrooms can serve as "learning labs" as well as struggling teachers in need of support.	• Models are not able to sort out teacher effects from classroom effects. • Vertical test alignment is assumed (i.e., tests essentially measure the same thing from grade to grade). • Value-added scores are not useful for formative purposes because teachers learn nothing about how their practices contributed to (or impeded) student learning. • Value-added measures are controversial because they measure *only* teachers' contributions to student achievement gains on standardized tests.

Classroom Observations

Description

Teacher observations take many forms, measure different aspects of teaching, and vary greatly in their implementation. They may be a district-developed set of categories that are used to give teachers' formative feedback. They may be a product purchased from an outside vendor that comes with rater training and scoring. Most often, observations occur somewhere between once and a few times during the school year, encompass roughly one lesson, and happen on a day agreed upon by the teacher and the rater. There is often a preobservation or postobservation conference between the rater and the teacher. The degree to which observations can or should be used for specific purposes depends on the instrument, how that instrument was developed, the level of training and monitoring raters receive, and the psychometric properties of the instrument. Review of the research suggests that observation scores have been related to important outcome measures such as student achievement (Gallagher, 2004; Kimball, White, Milanowski, & Borman, 2004; Milanowski, 2004).

When measuring teacher effectiveness through classroom observations, valid and appropriate instruments are crucial as well as trained raters to utilize those instruments in standard ways so that results will be comparable across classrooms. The following example may help explain what is meant by a "trained rater":

Presume that there are four aspects of teaching effectiveness one wants to measure: teacher student interactions, classroom management, school community contributions, and subject matter knowledge. Each is measured on a three-point scale: *needs improvement*, *satisfactory*, and *excellent*. In rater training, raters would be taught the differences among *needs improvement*, *satisfactory*, and *excellent* classroom management. What, for example, causes a specific classroom management technique to go from satisfactory to excellent? Raters would need to practice applying those criteria to a number of lessons to make sure they understand when they actually are faced with diverse actions. Raters also would be taught where particular actions— say, scaffolding students' understanding of fractions—are to be scored. In that example, raters might want to have such scaffolding fall into the interactions domain, whereas, others might tend to score scaffolding as a part of the teacher's subject matter knowledge. As a part of the training, these issues would be discussed and practiced, and hopefully raters would learn to score observations accurately against the standards. If this were to happen, the raters would be calibrated to the standards. If raters could do this consistently for numerous lessons, they would be reliably trained.

Whoever is using the raters' scores also would want to be sure that throughout the school year, raters are consistently applying those criteria. It would be problematic if scores were more lenient in the beginning of the year (because, for example, the teachers are just getting started) and more stringent in the middle or end of the year (because raters had seen a lot of teaching). This would mean one's scores would partially depend on when they were observed. In addition to issues of what day during the year observations take place, users of observation protocols also should pay attention to whether or not there is information (and training) to help raters consistently apply the rating criteria across different times of the day and subject matter.

Depending on the protocol, trainers may or may not have investigated, thought about, or developed training materials to deal with these issues. These issues are critical for any protocol, but they are especially important if scores are going to be used for high-stakes purposes such as tenure and compensation.

Given those technical considerations, observations can provide important, useful information about a teacher's practice if used thoughtfully. Districts must be careful, however, because observations are susceptible to rater biases in ways that some of the other measures of teacher effectiveness are not.

Examples

Examples of observation protocols that are widely used and have been studied on a relatively large scale include Charlotte Danielson's (1996) *Enhancing Professional Practice: Framework for Teaching* and the University of Virginia's Classroom Assessment Scoring System (CLASS) for prekindergarten and K–5 (Pianta, La Paro, & Hamre, 2006). The *Framework for Teaching* is meant to be used across subject matter and grade levels. CLASS also can be used across subject matter but has particular grade spans (early childhood, K–5, and 6–12).

In addition to these instruments, there are countless numbers of additional observation protocols that are less widely used, some of which have no published validity information and others of which have been used in very limited contexts—most often in research projects in which scores are not reported to teachers or used for any purpose outside the research project. A subset of these more narrowly used instruments is comprised of several promising subject-specific protocols. These protocols are particularly noteworthy, given the increasing focus on the role of subject-specific knowledge for teaching and the increasing call for teachers to have more and more relevant subject matter knowledge. Examples of these include the Reformed Teaching Observation Protocol (RTOP) for mathematics and science (Piburn & Sawada, 2000), the Quality of Mathematics in Instruction (QMI) in mathematics (Blunk, 2007), and the TEX-IN3 for literacy (Hoffman, Sailors, Duffy, & Beretvas, 2004). Though these three specific instruments are regarded as promising, they have not been widely used by anyone beyond the developers, and there is little published data on how these instruments function. RTOP has the most information (e.g., MacIsaac, Sawada, & Falconer, 2001; Piburn & Sawada, 2000; Sawada et al., 2002), whereas QMI is the newest and is still in the beginning stages of documentation (e.g., Blunk, 2007). For practitioners interested in modifying generic protocols to include more subject matter, these would be excellent resources. They also might be useful for districts interested in using subject-specific protocols for formative feedback.

Danielson's Framework. Danielson's (1996) *Framework for Teaching* is one of the most commonly used observation protocols in districts (Brandt, Mathers, Oliva, Brown-Sims, & Hess, 2007). Danielson based the framework on research she and colleagues conducted in developing Praxis III, an observational protocol designed by ETS for assessing the classroom performance of beginning teachers. ETS researchers worked with many teachers and other educators to do the following:

- Define a holistic view of teaching.
- Describe the complex relationships of teachers and students.

- Examine the importance of tailoring teaching to the individual, developmental, and cultural differences of students.

- Consider the influence of the subject being taught on teaching.

- Spell out the implications of all this for teacher assessment.

The *Framework for Teaching* is described on the Danielson Group website as "a research-based set of components of instruction, aligned to the INTASC standards, and grounded in a constructivist view of learning and teaching." It consists of four domains, broken down into 22 components and 76 smaller elements. Teachers are evaluated against a detailed rubric, which can be used to rate each of the 76 elements as *unsatisfactory*, *basic*, *proficient*, or *distinguished*. The framework can be used for several purposes, such as reflection and self-assessment, mentoring and induction, peer coaching, and supervision. Although it can be used for summative evaluation, providing feedback for formative use is key. According to the Danielson Group website:

> The *Framework* may be used for many purposes, but its full value is realized as the foundation for professional conversations among practitioners as they seek to enhance their skill in the complex task of teaching. The *Framework* may be used as the foundation of a school or district's mentoring, coaching, professional development, and teacher evaluation processes, thus linking all those activities together and helping teachers become more thoughtful practitioners.

The *Framework for Teaching* has been implemented and studied in districts including Cincinnati, Ohio; Reno/Sparks, Nevada; Coventry, Rhode Island; and Los Angeles, California; with several studies finding that teachers who scored higher on the *Framework for Teaching* were associated with greater gains in student achievement (Gallagher, 2004; Kimball et al., 2004; Milanowski, 2004; Milanowski, Kimball, & Odden, 2005). These findings vary by subject matter (reading and mathematics) and grade level and are small to modest sized correlations. It is important to note that there was wide variation in rater training, rater's relationship with the teacher (peer, supervisor, etc.), the degree of adherence to Danielson's recommendations for use, the use of the scores, and the number of observations conducted for each teacher. This variation may be partially responsible for the range of findings.

For example, the school in Los Angeles that used the *Framework for Teaching* adopted a subject specific version, used it for simultaneous formative and summative feedback, and linked scores with skills-based merit pay. In Cincinnati, a nonsubject-specific version was used for both formative and summative purposes and was linked to skills-based merit pay. Research reports on these sites reported observations taking place between three and six times per year by a hired teacher evaluator (who was released from teaching duties for three years) and/or an administrator. In Nevada, principal and assistant principals used multiple sources of evidence to assign scores on a nonsubject-specific version of the modified framework. This information was used formatively and summatively, but it was not used for compensation. In Coventry, Rhode Island, principals and department heads conducted observations based on a modified version of the *Framework for Teaching*. Frequency of observations was dependent on tenure status, and scores were not intended to be linked to pay. In the two cases in which *Framework for Teaching*

scores were used for compensation decisions, they were used with other information (e.g., credentials, experience).

This variation and the research documentation of the instrument suggest a number of important points. First, a good proportion of teachers in each site find the framework credible and helpful for their teaching (Heneman, Milanowski, Kimball, & Odden, 2006). Scores have been used in four districts and 179 schools across the country for both formative and summative purposes, which suggests it is possible to use the *Framework for Teaching* in various contexts and purposes. The framework is general with respect to grade level and subject matter area. It does not capture subject specific aspects of teaching, though at least one district was able to develop subject-specific versions. The research does not indicate whether modified versions of the instrument perform as well as versions that adhere to Danielson's recommendations. In addition, it is not evident whether the instrument functions differently (or is implemented differently) at different grade levels. Finally, the *Framework for Teaching* values a constructivist approach to teaching.

Classroom Assessment Scoring System (CLASS). This observation instrument was developed at the University of Virginia as a measure of classroom quality in preschool and in the early elementary grades. A number of studies have been conducted to examine the relationship between scores on CLASS and students' academic and social growth, as described in this section. CLASS was conceptually based on theories of child development, and the dimensions characterize interactions between students and teachers (Pianta, La Paro, & Hamre, 2007). In CLASS, "the focus is on what teachers *do* with the materials they have and in the interactions they have with students" (Pianta, La Paro, et al., 2007, p. 1). Although the instrument started out as a measure of classrooms in early elementary settings, protocols have now been developed for prekindergarten, Grades K–5, and Grades 6–12.

The CLASS framework is a theoretically driven and empirically supported conceptualization of classroom interactions organized into three major domains: emotional support, classroom organization, and instructional support. Each domain has a set of more specific dimensions of classroom interactions that are deemed to be important to students' academic and social development. The *emotional support* construct refers to the teacher's ability to establish a classroom climate and set of relationships that enhance students' social and emotional functioning. The *classroom organization* construct refers to classroom processes related to the organization and management of students' behavior, time, and attention in the classroom. The *instructional support* construct refers to teaching that is consistent with both theories of how students learn best and domain-specific models of content.

CLASS uses time-sampling in the form of observation cycles. A cycle is defined as a 30-minute period in which the first 20 minutes is used for observations and note-taking and the next 10 minutes are used for scoring. CLASS has been used both in real-time observations and videotaped lessons. The authors of CLASS found that "four cycles provides a representative sampling of classrooms" (Pianta, La Paro, et al., 2007, p. 10). Based on two large studies using CLASS, researchers also found that scores are relatively stable across the school year. There are, however, small differences in mean scores around the holidays and toward the end of the year.

The developers of CLASS offer training to groups interested in using the protocol. Training consists of a two-day training and scoring session in which potential raters watch numerous 20-minute training videos that have been consensus-scored by at least three master raters. At the end of the training, potential raters take a reliability test on five 20-minute segments of videotaped teaching. A rater is considered to have achieved sufficient reliability if he or she produces a score within one point of the master raters' consensus score for that video clip. The training materials thus far have been successful, achieving an average inter-rater reliability of 87 percent (Pianta, La Paro, et al., 2007).

Currently, there is little information on the Grades 6–12 version of CLASS works, but there is extensive validity and reliability data on the elementary and prekindergarten versions, and those data are promising (Pianta, La Paro, et al., 2007). The data on the prekindergarten and K–5 versions come from six studies in more than 1,700 PK–5 classrooms in urban, rural, and suburban settings across the country. Scores on CLASS or its precursor have been related to academic gains, other developmental markers, and student behavior (Hamre & Pianta, 2005; Howes et al., 2008; Rimm-Kaufman, La Paro, Downer, & Pianta, 2005).

Although the information provided suggests that the prekindergarten and K–5 instruments are of high-quality, there are a number of considerations to keep in mind:

- There is little information about the secondary instrument, and thus it should be used with caution.

- The protocol can be used across subject matters, but it is targeted at grade levels. The protocol does have an instructional support domain but is limited in terms of the kind of subject-specific information it can generate for formative purposes.

- There are increasing numbers of districts and schools using the protocol; however, the research does not reveal whether or how districts adapt or use the instrument.

In addition, it is not known whether districts find it affordable or doable to keep raters trained at reliable and calibrated levels. Many researchers find the scores from CLASS to be meaningful, but again, there is not much information about how teachers view CLASS scores.

Strengths and Cautions

As a class of instruments, observation protocols have a number of strengths. Teacher observations often *seem* valid. To the degree that observational ratings reflect who teachers and administrators believe is a good teacher, stakeholders can support their use. This makes it particularly important for a given protocol to be developed to reflect stakeholders' ideas about best practice and to be implemented in robust, defensible ways. When observation protocols clash with stakeholders' beliefs and/or are implemented in biased ways, the validity of results is weakened. Thus, including stakeholders' views about the content and implementation of observation protocols may be beneficial.

Another strength is that observation protocols have been and could be used as a part of teacher compensation. They have been modestly to moderately linked to student achievement, depending on the instrument. They also have been used both formatively and summatively, suggesting that the same instrument can serve multiple purposes for districts. For formative use, observations can provide rich feedback about teachers' areas of strengths and weaknesses. This type of rich feedback could be used productively for formative evaluations of teachers. The rater/evaluator can share with the teacher the results of the evaluation and then use those results to help develop (cooperatively) a plan of professional development and personal growth that will lead to a closer alignment to the effective teaching practices that are valued.

There are a number of cautions that are worth bearing in mind, however, considering the use of observations for evaluation of any form. The most popular and well-researched instruments are generic and may not take account of subject-specificity in ways that could support teachers as they endeavor to teach more students increasingly ambitious content. Many protocols have been used in research projects only by the researchers themselves (or by one other researcher who was not involved in the protocol's development). This lack of field testing introduces two significant concerns. First, for many instruments, it is not evident whether it is possible for districts to use the protocols effectively for nonresearch purposes. This issue might be resolved by a review of the instruments themselves and a conversation with the developers, but nonetheless, it is important to note there is little research to guide practitioners on this issue. In addition, because many protocols have not been used to improve practice, it is not known whether the district can expect to see a change in teachers' practice when a particular protocol is used. This is a serious gap in the understanding of how these protocols might improve practice.

In addition, the link between observations and student achievement and other outcome measures (e.g., graduation and citizenship) is another concern. Though there have been some studies that link teachers' scores on observation protocols to gains in student achievement (Gallagher, 2004; Kimball et al., 2004; A. Milanowski, 2004), there is much work to be done. For example, there is little research that links scores on well-validated observation protocols with other student outcomes of interest. Observations teachers may tell a great deal about how well a given teacher's practice aligns with what is believed to be good practice, but without linking this information to student outcomes, determining effectiveness is difficult.

A final set of concerns about observation protocols involves the issue of raters. Proper training is essential because raters are making moment-by-moment judgments about what they see. McGreal (1990) contends, "The high inference nature of rating scales places the burden of selecting a rating directly upon the evaluator" (p. 50). Considerable progress has been made in developing methods for ensuring more consistent ratings through evaluator training and calibration sessions. However, there is no assurance that a given state or district actually employs these methods, meaning that different evaluators might give very different scores to the same teacher, depending on their views of good teaching. Measuring teacher effectiveness through observations can be very uneven, which threatens the utility and credibility of the protocols themselves.

Principal Evaluations

Description

Classroom observation conducted by principals or vice-principals is one of the most common forms of teacher evaluation (Brandt et al., 2007). The format varies by district; for instance, a principal evaluation can consist of a formal observation using a validated instrument, conducted at a predetermined time, coupled with pre-interviews and post-interviews with teachers, and used for both formative and summative purposes (Heneman, Milanowski, et al., 2006). It also can be an informal drop-in visit by the principal, used to develop a quick impression of how and what a teacher is doing in the classroom.

Principal evaluations differ from evaluations performed by district personnel, researchers, or other outside evaluators who are hired and trained to conduct evaluations. Principals are most knowledgeable about the context of their schools and their student and teacher populations, and thus may be likely to compare the school's teachers to each other rather than to the larger population of teachers in the district or state. They may employ evaluation techniques that serve multiple purposes:

- To provide summative evaluation scores for school, district, state, or federal accountability purposes.

- To inform decisions about tenure or dismissal.

- To identify teachers in need of remediation.

- To provide formative feedback to improve teachers' practice.

Although these factors can make principals valuable sources of information about their schools and teachers, they also have the potential to introduce bias in either direction to principals' interpretation of teaching behaviors.

Examples

Although principal evaluation is the most common component of teacher evaluation systems, there is not a lot of solid evidence on the validity of these evaluations. One recent study by Brandt and colleagues (2007) examined district policies on teacher evaluation in several Midwestern districts. They found that principals and administrators typically conducted the evaluations, which were primarily focused on making decisions about which beginning teachers should be retained and released. District policies were more likely to offer guidance on the process of conducting evaluations than to instruct administrators on the potential uses of the evaluation results. Two particularly relevant findings from the study are that most evaluations were summative—for high-stakes employment decisions, rather than formative—for helping teachers grow in the profession. Furthermore, only 8 percent of districts mentioned evaluator training as a component of their teacher evaluation systems. Thus, although the use of high-stakes, summative assessment was prevalent, the evidence that assessments were used in a reliable and valid manner was not. These findings may be regional rather than national; however, they raise the concern that career consequences are being based on the assessments of evaluators who may have little understanding of how to use the instrument in ways that ensure valid scores.

Other studies have examined the accuracy and predictive value of principal evaluations by comparing subjective principal ratings of teachers to value-added scores of student achievement (Harris & Sass, 2007b; Jacob & Lefgren, 2005, 2008; Medley & Coker, 1987; Wilkerson, Manatt, Rogers, & Maughan, 2000). These studies required principals to rate teachers in their school using a scale created by the researcher. Because these ratings were not based on a specific observation and were not tied to any official decision making, these studies are distinct from the context of principal evaluation as it generally occurs in schools, but they do raise noteworthy issues about the accuracy of principals' judgments. The main finding from these studies is that principal ratings are significantly correlated with teacher value-added scores, but the correlation is usually low. Principals were found to be fairly accurate at identifying teachers in the top or bottom group of effectiveness but were less successful at distinguishing between teachers in the middle (Jacob & Lefgren, 2008). Note, however, that the same result has been found for value-added measures (e.g., Archibald, 2007; McCaffrey et al., 2003). Principals were better able to predict value-added scores at the elementary level than they were at the secondary level (Jacob & Lefgren, 2008) and were better at making reasonable judgments about which teachers would improve achievement in mathematics than they were in making judgments about which teachers would improve achievement in reading (Harris & Sass, 2007b; Wilkerson et al., 2000).

Findings do indicate that principal ratings are better predictors of teacher value-added scores than several standard measures of teacher quality (e.g., experience, certification, and education) (Harris & Sass, 2007b); however, some of the specific findings present a mixed picture. Harris and Sass (2007b) found that principal ratings were as accurate at predicting future student achievement gains as value-added measures of teacher effectiveness, whereas Jacob and Lefgren (2008) found principal ratings to be less accurate predictors than value-added measures. Wilkerson and colleagues (2000) found that student ratings of teachers were better predictors of achievement than principal ratings. Jacob and Lefgren (2008) also explored some of the speculations behind why the correlation between principal ratings and value-added scores was lower than expected and found that principals may tend to pay more attention to the mean level of achievement in a teacher's class and not the relative improvement that students made (i.e., they do not account for differences in classroom composition). In addition, they found that principals may tend to focus on their most recent observations of a teacher rather than considering the teacher's long-term performance. Their data support the notion that a combination of principal ratings and value-added measures is a stronger predictor of student achievement than either alone.

Strengths and Cautions

Given the many areas a principal must attend to simultaneously and in the interest of reducing the subjectivity and potential bias inherent in observation, it is advisable for administrators to employ a specific and validated observation protocol when conducting teacher evaluations (see the Classroom Observations section on page 20 for examples), especially if the information is to be used in any high-stakes decision making. When choosing an instrument, careful attention should be paid to its intended and validated use. As discussed in the observation section, administrators should be fully trained on the instrument, rater reliability should be established, and periodic recalibration should occur.

Observations should be conducted several times per year to ensure reliability, and a combination of announced and unannounced visits may be preferable to ensure that observations capture a more complete picture of the teacher's practices. Another consideration is the focus of the evaluation. For instance, an observation assessing deep or specific content knowledge may be better conducted by a peer teacher or content expert, as a principal or administrator may not be equipped with the specialized knowledge to make the best judgments necessary for this type of evaluation (Stodolsky, 1990; Weber, 1987; Yon, Burnap, & Kohut, 2002). Using a combination of principal and peer raters is another consideration that may increase the credibility of the evaluation.

To incorporate all of these ideas, principals should consider a *system* of evaluation that serves both formative and summative purposes and involves teachers in the process. If principals are viewed as uninformed or unjust evaluators, teachers may in turn not take evaluation procedures seriously. Making teachers aware of the criteria against which they are being judged ahead of time, providing them with feedback afterward, giving them the opportunity to discuss their evaluation, and offering them support to target the areas in which they need improvement are all components that will strengthen the credibility of the evaluation. Evaluation systems are more likely to be productive and respected by teachers if the processes are explained well and understood by teachers, well-aligned with school goals and standards, used formatively to inform teaching and encourage professional development, and viewed as a support system for promoting schoolwide improvement.

Analysis of Classroom Artifacts

Description

Another method that has been introduced to the area of teacher evaluation is the analysis of classroom artifacts, such as lesson plans, teacher assignments, assessments, scoring rubrics, and student work. The classroom artifacts that a teacher selects and creates and the student work that is generated can provide insight into the types of opportunities to learn that students are presented with on a day-to-day basis. Depending on the goals and priorities of the evaluation, artifacts may be judged on a wide variety of criteria including rigor, authenticity, intellectual demand, alignment to standards, clarity, and comprehensiveness. Though the examination of teacher lesson plans or student work is often mentioned as a part of teacher evaluation procedures, few systems employ a structured and validated protocol for analyzing artifacts to evaluate the quality of instruction. Use of a valid protocol for analyzing teacher assignments and student work introduces a potentially useful compromise in terms of providing a window into actual classroom practice, as evidenced by classroom artifacts, while employing a method that is less labor-intensive and costly than full classroom observation.

Examples

Instructional Quality Assessment (IQA). The most work on this has been done by the National Center for Research on Evaluation, Standards, and Student Testing (CRESST) located at the University of California–Los Angeles. CRESST researchers have worked extensively to develop the Instructional Quality Assessment (IQA), a protocol that can be used both for evaluating the instructional quality of a classroom and for providing feedback to teachers for purposes of professional development. IQA consists of protocols for rating the quality of teachers' assignments and student work in reading comprehension and mathematics. Rubrics focus on quality of discussion, rigor of lesson activities and assignments, and quality of expectations communicated to students (Matsumura, Slater, Junker et al., 2006). CRESST has conducted several pilot studies on IQA, finding that the rubrics are generally correlated with quality of observed instruction, quality of student work, and standardized student test scores (Clare & Aschbacher, 2001; Junker et al., 2006; Matsumura, Garnier, Pascal, & Valdés, 2002; Matsumura & Pascal, 2003; Matsumura, Slater, Junker et al., 2006). These studies also indicate reasonable reliability for the instrument, though more work may be needed to confirm its dependability and stability. For instance, work has been conducted to determine the ideal number of assignments that should be collected to maximize accuracy of scores while minimizing teacher time and effort.

Intellectual Demand Assignment Protocol (IDAP). Newmann and colleagues of the Consortium on Chicago School Research have conducted another branch of work on analyzing instructional artifacts (Newmann et al., 2001; Newmann, Lopez, & Bryk, 1998). These researchers were interested in determining the authenticity and intellectual demand of classroom assignments and created rubrics for scoring teacher assignments and student work in mathematics and reading. The rubric assesses the degree to which the assignment involves construction of knowledge, promotes disciplined inquiry, and exhibits value beyond school. The authors collected "typical" and "challenging" assignments from Chicago elementary school teachers, which were rated by trained scorers according to the rubric (see Newmann et al., 1998). Scorers were able to achieve high levels of interrater reliability using the rubrics, with greater than 90 percent agreement within one point for the different subjects and grades scored. IDAP scores were matched to student achievement gains in each teacher's classroom. Findings showed that in classrooms with higher-scoring assignments, student learning gains on the Iowa Test of Basic Skills were 20 percent higher than the national average; in classrooms with lower-scoring assignments, learning gains were 22 percent to 25 percent lower than the national average. Use of high-demand assignments appeared unrelated to student demographics and prior achievement and benefited students with high and low prior achievement alike.

Scoop Notebook. Another example is the Scoop Notebook—developed and piloted by Borko, Stecher, Alonzo, Moncure, and McClam (2005) and further analyzed by Borko, Stecher, and Kufner (2007)—used to evaluate classroom practices through the examination of artifacts reflecting the teaching and learning process. Materials in the notebook included handouts, scoring rubrics, writing on the board, student class work, student homework, and projects. In a pilot study of 13 middle-school mathematics and science teachers, teachers provided two examples of "high" and "average" quality work for each set of class work or homework collected over a five- to seven-day period. Teachers also took pictures of artifacts in the classroom (e.g.,

writing on the board) and answered reflective questions about lessons. Multidimensional scoring guides were developed by the researchers using mathematics and science education standards and were rated by two or more trained raters. Although rater agreement was higher than would be predicted by chance, there were clear areas in which raters were inconsistent, and they appeared to be better at judging a lack of evidence rather than the presence of evidence. Some teachers found the process to be beneficial to their instruction, particularly reflecting on the lessons. Ratings also were found to be reasonably consistent with observational measures, but no links were made to student achievement in this small pilot study.

Strengths and Cautions

Analysis of classroom artifacts is a promising method to provide a comprehensive view of a teacher's quality of instruction and gain a deeper understanding of his or her intentions and expectations. It may prove to be a practical and feasible method, as the artifacts have already been created by the teacher and the procedures do not appear to place unreasonable burdens on teachers (Borko et al., 2005). This method has the potential to provide summative information about instruction as well as rich formative information and opportunity for reflection to teachers.

However, several cautions should be taken into consideration. As with the other methods discussed so far, accurate scoring is essential to preserving the validity of the instruments. This requires adequate training and calibration of scorers and also may require scorers to possess some knowledge of the subject matter being evaluated. Some studies also have noted that a lack of variation in quality of assignments (i.e., teachers at a school consistently assign very low-quality assignments) can make it difficult to validate the scoring rubrics (e.g., Matsumura, Patthey-Chavez, Valdés, & Garnier, 2002). More research needs to be done to investigate the reliability and stability of ratings and explore links to student achievement. There remains a lack of research documenting the use of these instruments in practice, and they have yet to be validated by independent research efforts. Thus, much more work is needed to validate the use of this method in actual evaluation settings before it should be considered as a primary means for teacher evaluation.

Portfolios

Description

Portfolios are a collection of materials compiled by teachers to exhibit evidence of their teaching practices, school activities, and student progress. They are distinct from analyses of instructional artifacts in that portfolio materials are collected and created by the teacher for the purpose of evaluation and are meant to exhibit exemplary work, as opposed to a sampling of artifacts that are already being used in a teacher's classroom. The materials gathered are intended to demonstrate fulfillment of certain predetermined standards, and often portfolios are designed to promote teacher reflection and improvement in addition to being used for evaluation. Examples of portfolio materials include teacher lesson plans, schedules, assignments, assessments, student work samples, videos of classroom instruction and interaction, reflective writings, notes from parents, and special awards or recognitions. Part of the exercise for teachers is choosing a feasible number of artifacts that will represent the full range of their teaching practices and larger

school contributions while demonstrating how their performance meets the given standards. The portfolio process often requires a defense of why artifacts were included and how they relate to the standards (Painter, 2001).

Portfolios are commonly used in teacher preparation programs as a requirement for licensure, but states have increasingly adopted portfolio assessments for use in evaluating both beginning and experienced teachers. Vermont reformed their performance assessment program beginning in 1988, implementing a unique system that used performance assessments, namely portfolios, as a main source of evaluation instead of an addition to a more traditional program (Koretz, Stecher, Klein, & McCaffrey, 1994). Connecticut also has a well-known program, the Beginning Educator Support and Training (BEST) program, which requires teachers to complete portfolios as part of their continuing licensure requirements. Washington State's Professional Certificate Program offers an advanced certification that requires the completion of a classroom-based portfolio (see Office of Superintendent of Public Instruction, n.d.), and the state of Wisconsin has a voluntary Master Educator License that requires a teacher to demonstrate advanced proficiency on a portfolio assessment aligned with the Wisconsin Educator Development and Licensure Standards (see Wisconsin Department of Public Instruction, 2008). To illustrate the uses of portfolios in evaluation, Connecticut's BEST program and the well-known advanced certification program of the National Board for Professional Teaching Standard (NBPTS) are described in the following section.

Examples

Connecticut's Beginning Educator Support and Training (BEST) Program. The BEST program is a two-year induction, support, and assessment program for new teachers in the state of Connecticut. The first year consists of seminars, workshops, and meetings with an assigned mentor teacher, giving new teachers an opportunity to develop their practice. During the second year, teachers submit a portfolio for assessment of their practice, and a satisfactory evaluation is required for teachers to obtain full certification and remain teaching in the state. Teachers who do not pass the assessment must undergo further professional development and resubmit the portfolio during the third year; if they do not pass in the third year, they are no longer permitted to teach in Connecticut public schools. As a part of the program, teachers are entitled to school-based support in the form of mentorship, release time, and content-specific instructional support and to state-based support in the form of professional development seminars, conferences, and Internet-based resources. In turn, beginning teachers are expected to fulfill the requirements of the BEST program and keep their certification up to date using the resources provided to them (Connecticut State Department of Education, 2007; Pecheone & Stansbury, 1996).

The evaluation standards for BEST portfolios are culled from Connecticut's Common Core of Teaching standards and are based on demonstrating foundational skills that are believed to be common across teachers in all grade levels and subjects as well as establishing knowledge and competency in discipline-specific areas. BEST portfolios include "daily lesson plans for a five- to eight-hour unit of instruction with one class; two to four videotaped segments of teaching equaling in total approximately 30–40 minutes; examples of the work of two students; and reflective commentaries on teaching and learning that took place during the unit" (Connecticut State Department of Education, 2007, p.. 22). Portfolios are scored by experienced teachers in

the same discipline as the teacher being evaluated. These assessors are hired by the Connecticut State Department of Education, work for two years at the department as teachers in residence, and must participate in at least 50 hours of comprehensive training in scoring and pass reliability assessments. After portfolios are scored, teachers are provided with an individualized performance summary, which discusses their performance according to the categories of designing and implementing instruction, assessment of learning, and analysis of teaching. Portfolios are scored based on a series of discipline-specific guiding questions and performance indicators, which are included in portfolio handbooks so that teachers are fully aware of the evaluation criteria as they create their portfolios (Connecticut State Department of Education, 2007).

National Board for Professional Teaching Standards (NBPTS) Certification. NBPTS offers a certification system to recognize accomplished teachers who meet high and rigorous standards, and a main component of their evaluation is a portfolio assessment (the other component is an assessment of subject matter knowledge). NBPTS offers 25 certificates that cover a variety of subject areas and student developmental levels. Standards for certification in each area are created by committees of expert teachers and specialists in education, child development, and other relevant areas. The portfolio requirement consists of four different entries, three of which are classroom based and one which exhibits work with families, the community, colleagues, and the larger profession. Contents of the portfolios include video of instructional practice, video of teacher-student interactions, and student work samples; all entries must be accompanied by detailed reflection and analysis of the instruction represented. Portfolios are evaluated by assessors who have completed intensive training through NBPTS and met qualification requirements by demonstrating an understanding of the NBPTS standards, directions, scoring guides, and rubrics. Teachers and school counselors, especially those who have achieved National Board Certification, are eligible to apply to become assessors (National Board for Professional Teaching Standards, 2008).

Much research has been conducted on NBPTS certification. There are several studies linking NBPTS certification to gains in student achievement (e.g., Cavalluzzo, 2004; Clotfelter, Ladd, & Vigdor, 2006; Goldhaber & Anthony, 2004; Vandevoort et al., 2004), though there are also studies that do not find a relationship (e.g., Cunningham & Stone, 2005; McColskey et al., 2005; Sanders, Ashton, & Wright, 2005). In a recent evaluation commissioned by the U.S. Department of Education on the effects of NBPTS certification, the Committee on Evaluation of Teacher Certification determined that NBPTS certification is successful in identifying high-performing teachers, but not enough evidence exists to determine whether the process itself leads to improvements in practice or whether teachers who are already effective complete the process (Hakel, Koenig, & Elliott, 2008). Because NBPTS participation is strictly voluntary, findings from studies examining the impact of the NBPTS process on teachers can be hard to interpret. Teachers who pursue the certification are a self-selected group and may differ in significant ways from the teaching population as a whole (Pecheone, Pigg, Chung, & Souviney, 2005). Though the NBPTS process tends to be viewed by teachers as contributing to their learning and professional growth, these findings are based mainly on teacher or administrator perceptions (Pecheone et al., 2005) and have not yet been verified by studies using more direct measures of learning (Hakel et al., 2008).

Validity and Reliability Research. Portfolios can offer a very comprehensive and in-depth portrait of teaching activities; however, their complexity can raise concerns about the ability of scorers to evaluate them reliably. In a study on the implementation of the Vermont teacher assessment program, Koretz et al. (1994) discuss problems with the portfolio rating system in establishing rater reliability and distinguishing real differences in the quality of student work contained in the portfolios. They also describe related difficulties with establishing validity of the measure and using it for school accountability purposes.

Johnson, McDaniel, and Willeke (2000) point out that studies that have examined the interrater reliability of large-scale portfolio assessments have found that the percentage of agreement is usually between 45 percent to 75 percent, and correlations between raters rarely reach .80, which is considered by some as a necessary threshold of reliability. [The study cites Nunnally's (1978) argument "that test reliability of .80 was necessary for review of group means and at least .90 was necessary for reporting individual scores" (Johnson et al., 2000, p. 74)]. Thus, although some of these correlations are moderately high, they are lower than desirable for use in any high-stakes decision making. Johnson et al. also demonstrate that reliability is affected by the type and number of items being scored. In their investigation of interrater reliability for a smaller-scale family portfolio assessment, they examine separately the interrater reliability of ratings on six individual criteria, the composite of those six ratings, and an overall holistic rating. They found that in general, the reliability of rating individual criteria was consistently lower than the composite score and somewhat lower than the holistic score. They also conducted a decision study to determine the number of raters necessary to achieve a reasonable level of reliability for each of these categories, finding that three raters were desirable for the individual rating or the holistic rating but that two raters were sufficient for the composite rating.

Tucker, Stronge, Gareis, and Beers (2003) examined the validity and usefulness of teaching portfolios in assessing teacher performance for both accountability and professional development purposes. In teams of two, researchers rated a random stratified sample of 24 portfolios from elementary, middle, and secondary teachers, based on 18 teacher responsibilities specified by the district covering four major domains (instruction, assessment, management, and professionalism). Perceptions of the usefulness of portfolios were measured via survey and follow-up focus groups with teachers and administrators. Authors found that portfolios were able to document the fulfillment of the 18 teaching responsibilities and included representation of each of the four major domains, and 90 percent of the artifacts submitted had content validity (i.e., were relevant to the domains). Professionalism was the most highly represented domain, illustrating the role of portfolios in documenting aspects of teacher performance that cannot be measured through classroom observation. Administrators found that portfolios gave them a broader view of teacher activities and allowed them to make "finer distinctions about the quality of teacher performance" (Tucker et al., 2003, p. 572). Both teachers and administrators viewed portfolios as fair and accurate, but teachers expressed concerns about feasibility. There were mixed results regarding the usefulness of portfolios for professional growth, with some teachers reporting them helpful for reflecting on practice but with little evidence of impact on teaching practices. Tucker et al. suggest that teachers may need further training in order to make the connection between teaching reflections and changes in instructional practice.

Overall, these studies illustrate that although portfolios are an effective method for tapping into broader concepts of teacher development and providing valuable information to teachers about their practice, several issues in scoring portfolios exist, and more research is needed to fully assess their reliability and validity. Due to these concerns, some studies advise against the use of portfolios as a stand-alone assessment in high-stakes decision making (e.g., Johnson et al., 2000). In addition, there is a lack of studies that investigate the relationship between scores on portfolio assessments and student outcomes, and this area deserves much more research.

Strengths and Cautions

Portfolios do offer several advantages over some of the other measures of evaluation discussed. They are generally considered useful for providing a broad and varied view of a teacher's many capabilities and providing formative information and opportunities for teacher reflection that can enhance performance. They can be used with teachers in any subject or grade level and thus are useful in multiple contexts. They are a very comprehensive measure, with the ability to assess aspects of teaching that are not readily observable in the classroom and extend beyond classroom instruction. They also have high face validity, generally being viewed by teachers and administrators as "authentic" assessments that are relevant and useful to their teaching practice. Portfolio assessments provide the opportunity to actively involve teachers in the evaluation process and give them personal ownership of their improvement and professional growth, helping to reform the conception of evaluation as something done *to* teachers *by* administrators (Tucker et al., 2003).

As this discussion indicates, more research on the reliability and validity of portfolios as a performance assessment is needed before they should play a substantial role in evaluation for accountability purposes. They present a useful opportunity for providing formative assessment to teachers, though teachers may need training in order to learn how to choose relevant artifacts (Painter, 2001) and reflect on their practice in a way that fosters improvement and leads to actual changes in practice (Tucker et al., 2003). They also can become quite cumbersome for teachers, requiring a significant time commitment if they are to gain the most benefit from the portfolio process, thus it is recommended that teachers are provided with support and time to complete portfolio requirements. In a study of beginning teacher performance assessments in California, Mitchell, Scott, Hendrick, and Boyns (1998) found that the amount of priority placed on the program by the school and district was related to teachers' perceptions of fairness and helpfulness of the assessments (cited in Pecheone et al., 2005). This demonstrates how buy-in and support from the administration can be crucial to the success of a performance assessment program.

Tucker et al. (2003) make some useful observations and suggestions based on their validity study. They recommend that to maximize the efficacy of portfolio assessments, it is useful to include complete units of study (e.g., lesson plans, teaching strategies, sample assessments, and scoring rubrics; student work with teacher comments that pertain to the specific unit; and reflections on the artifacts the teacher chose to include with explanations of their relevance and importance). They also recommend the use of portfolios inclusively but not exclusively in the evaluation of teachers, as a complement to data collected through classroom observation, conferences, and client surveys.

Self-Reports of Teacher Practice

Description

This section examines different categories of self-report measures of teacher performance. These measures prompt teachers to report on what they are doing in the classroom and may take the form of surveys, instructional logs, and interviews. These measures vary widely depending on the focus, the level of detail they attempt to gather, and the intended use of the scores. Mullens (1995) describes several considerations in reference to designing large-scale survey measures of teaching, such as whether or not the aspects measured bear a relationship to student achievement or other outcomes of interest, whether the measures can inform policy and decision making aimed at educational improvement, and whether the measures can be used appropriately with the population of interest. For instance, as discussed in the observation section, survey measures may focus on broad and overarching aspects of teaching that are thought to be important in all contexts, or they may focus on specific subject matter, content areas, grade levels, or techniques. Survey measures may consist of straightforward checklists of easily observable behaviors and practices; they may contain rating scales that attempt to assess the extent to which certain practices are used or aligned with certain standards; or they may set out to measure the precise frequency of use of practices or standards. Thus, this class of measures is quite broad in scope, and considerations in choosing or designing a self-report measure will depend largely on its intended purpose and use.

Examples

Surveys. Several large-scale and well-known teaching surveys focus on measuring reform-oriented practices or enactment of curriculum. Examples of large-scale surveys include those developed by the National Center for Education Statistics (NCES); the Trends in International Mathematics and Science Study (TIMSS); Reform-Up-Close and the *Surveys of Enacted Curriculum* (SEC); and studies by the RAND Corporation, including the School Reform Assessment Project, Validating National Curriculum Indicators, and the California Learning Assessment System (CLAS). Some of these are broad and meant to be used with all teachers (e.g., NCES survey), whereas others are subject-specific and focused on content (e.g., TIMSS and CLAS surveys). Mullens (1995) identifies four broad dimensions of classroom instruction that are critical for large-scale surveys to address: pedagogy, professional development, instructional materials and technology, and topical coverage within courses. According to Mullens (1995), "All four dimensions under consideration have an established or expected relationship to student achievement and could provide interesting information about variation in achievement. Of the four, pedagogy and topical coverage within courses are more related to the teacher/student interaction and may therefore have a stronger relationship with student achievement" (p. 16).

One example of a thoughtfully developed and tested large-scale survey is the SEC, which were created as practical and reliable tools for data collection and reporting on instructional practices and content being taught in K–12 mathematics, science, and English language arts (ELA) classes. Blank, Porter, and Smithson (2001) describe how SEC data can be used in schools. The survey is conducted online, so results are tabulated and made accessible to schools in a variety of

formats. Data from the SEC allow administrators to examine differences between schools and teachers, compare instruction to standards, and evaluate the alignment between practices and standards. Like any effective evaluation instrument, it also provides a framework for communicating about practices and instruction, which can guide teacher reflection and lead to increased discussion and collaboration among colleagues. Blank et al. (2001) address concerns in the study about potential inconsistencies or inaccuracies in teacher responses due to factors such as differing interpretations of the terminology used and the time lag in reporting (teachers reported on their practices for the entirety of the semester or year). They also address concerns about low response rates; however, they express confidence in the accuracy of the teacher reports, citing findings from an earlier related study of Reform-Up-Close (Porter, Kirst, Osthoff, & Smithson, 1993), which compared teacher practices as measured by daily logs, independent observation, and teacher survey reports and found survey data to be highly correlated with the more detailed and frequently collected log measures.

Other studies also have investigated the validity of self-report survey data by comparing multiple measures. A study conducted by RAND Corporation researchers examined teachers' instructional practices using both self-report survey data and analysis of artifacts from teachers' classroom activities (Burstein et al., 1995). Researchers collected homework, quizzes, classroom exercises, projects, and exams from 70 mathematics teachers in California and Washington. They also analyzed daily logs kept for five weeks by the participating teachers, which described their instructional practice. The researchers found problems with the validity of the survey responses, stating, "instructional goals cannot be validly measured through national surveys of teachers. The data are inconsistent not only with artifact data but also with teachers' own self-reports on other survey items such as those describing their exam formats" (Burstein et al., 1995, p. 54). This finding raises concerns about the use of self-report survey data to represent teacher practices. It also might suggest that evaluating classroom artifacts, while considerably more expensive, may provide better evidence of actual teacher practices than self-report data. However, more research is needed to examine the validity of these measures.

Mayer (1999) conducted a study to examine the validity of teacher self-report data on instructional practices by surveying Algebra I teachers on their use of practices that reflected teaching standards set forth by the National Council of Teachers of Mathematics (NCTM). The author calculated the time teachers reported spending on certain practices aligned with the standards, comparing this with observational measures of the time they spent engaging in those practices. The study found that observational and survey measures were highly correlated but that survey measures were systematically inflated. It also determined that measures of individual practices were not reliable; however, composite measures of teaching practices were valid, and relative rankings of practices used were generally consistent. In other words, the survey could indicate the extent to which a teacher utilized a group of instructional practices as compared to other teachers but could not accurately measure the amount of time spent on individual practices. In addition, when a teacher reported using certain practices, the survey did not reveal anything about the level or quality of their implementation. Though sample sizes were small, these findings reveal important distinctions about the quality of information that can be gleaned from self-report survey data.

Logs. In contrast to broad surveys, instructional logs require teachers to keep a frequent and detailed record of teaching. The logs are highly structured and ask for specific information regarding content coverage and use by both the teacher and students. Much of the development and research work in the area of instructional logs has been conducted by researchers from the Consortium for Policy Research in Education (CPRE), as part of their larger Study of Instructional Improvement. The study is a comprehensive examination of measures of teaching, using multiple methods to gather data on instruction, including questionnaires, instructional logs, classroom observations, and teacher interviews. Ball and Rowan (2004) describe how the logs came to be developed: "Because gathering annual data on daily instruction likely often misrepresented actual practice, more frequently administered logs emerged as an approach to gathering information about content covered" (p. 4).

Camburn and Barnes (2004) examined the validity of these instructional logs, focusing on language arts lessons, by comparing teacher log responses with responses given by third-party observers. The log consisted of 150 items, including detailed information on content and emphasis on curricular areas. Thirty-one teachers who were pilot-testing the logs in eight public elementary schools were observed for one day, and both the teachers and observers completed a log for each lesson. One of the main findings revealed that teacher and researcher reports did not always agree, and scores between researchers were nearly always more highly correlated than scores between researchers and teachers, indicating that "researchers and teachers may have brought different perspectives to bear when completing the language arts log, perhaps drawing on different knowledge and experiences" (p. 59). Authors speculated that because observers have a more limited experience with the classroom than teachers, they may lack certain contextual information or interpret information differently when making judgments that reflect how a teacher perceives his or her intentions and practices. The importance of establishing a common understanding of terminology between teachers and raters also was raised, as differing interpretations of glossary terms may have contributed to inconsistencies in ratings. The study also found that rater agreement was affected by the degree of detail in the category being scored, the frequency of the instructional activity, and the content being covered.

In addition, Camburn and Barnes (2004) suggest that the ability to create a clear shared understanding with teachers through a log remains a challenge and is a significant threat to construct validity. They argue that researchers may face a trade-off between measuring subtle differences in content use that may affect student learning and the use of categories that measure broader aspects of instruction. They explain that "the former approach, which parses instruction more finely, makes interrater agreement more difficult to obtain and poses a threat to the validity of the measures. The latter approach may miss nuances in instruction that are theoretically and empirically important but may yield more valid measurement" (pp. 65–66). This study raises an important issue, which relates to the aforementioned studies: discrepancies between teacher self-reports of practice and third-party observer reports may not simply reflect inaccuracy on the part of the teacher but may uncover a larger issue concerning the differing values, knowledge, and interpretations that these two parties inherently bring into their evaluations. This is certainly an area worthy of further investigation.

Interviews. Another method for investigating teachers' self-reported practices is to utilize an interview protocol. Interviews are most often used as supplements to other measures of teaching

and are particularly useful in providing qualitative information that supports or explains results obtained from more quantitative measures. Studies that attempt to triangulate several measures of teaching in order to ensure accuracy of the results may employ an interview protocol, such as the aforementioned Study of Instructional Improvement (see Ball & Rowan, 2004) and the RAND Mosaic Study (see Le et al., 2006). The Mosaic Study examined the use of reform-oriented teaching practices, employing several measures including teacher surveys, instructional logs, structured vignettes, and observers' ratings of classrooms. An interview protocol was developed to investigate whether teachers felt that local systemic reforms and other policies were influencing their practices. This illustrates the very unique role interviews can play in gathering information on perceptions and opinions that may inform the "whys" and "hows" of measuring teacher performance and its impact.

Interview protocols can be highly structured or largely open-ended and can be a means for gathering data on practice that is more detailed or in-depth than survey measures. They are generally locally designed and intended for use in the context for which they were created. Few studies examine the reliability or validity of interview protocols intended to be used on a larger scale. One example is a study by Flowers and Hancock (2003), which describes the development of an interview protocol focused on professional standards and student learning. They describe the advantage of their interview protocol as a "method of collecting data from multiple sources while avoiding the shortcomings of singularly focused evaluation systems" (p. 163). The interview questions require teachers to provide specific examples of their instructional activities, intentions behind the activities, and specific actions they have taken to monitor and improve student learning. The protocol includes a structured scoring rubric with detailed criteria included for each rating. Evaluators must be trained on the interview protocol and scoring rubric, and teachers should be provided with the interview procedure and standards prior to the interview so that they can prepare materials in advance and formulate any clarifying questions they may have. This study reports high interrater reliability and rater consistency for the protocol, and extensive feedback from experts in the field helped to establish its content validity.

Strengths and Cautions

Teacher self-report methods may be one useful element in a teacher evaluation system, as they do have certain advantages. Self-report data can tap into a teacher's intentions, thought processes, knowledge, and beliefs better than the other methods discussed, and they can be useful for teacher self-reflection and formative purposes. In addition, it is important to consider the perspectives of teachers and involve them in their own evaluation because they are the only ones with full knowledge of their abilities, classroom context, and curricular content, and thus can provide insight that an outside observer may not recognize. Surveys are a cost-efficient, generally unobtrusive way to gather a large array of data at once. Using one instrument, data can be collected on instructional practices as well as administrative support, professional development opportunities, relationships with students, school climate, working conditions, demographic or background information, and perceptions or opinions that may have bearing on the effectiveness of a teacher.

Teacher self-report measures may be an efficient means of obtaining information about instructional practices without incurring the high costs of observation or other measures and can

be particularly useful as a first step toward investigating some question of interest (e.g., establishing some basic level of standard use and understanding among teachers) (Cohen & Hill, 2000; Spector, 1994). However, extreme caution should be taken not to base potentially consequential decisions on results of self-report measures. Research findings on the reliability and validity of these methods have produced mixed results. Concerns have been raised in the literature about self-report responses being susceptible to social desirability, defined by Moorman and Podsakoff (1992) in the organizational psychology literature as "the tendency on the part of individuals to present themselves in a favourable light" (p. 132). This phenomenon would include both the conscious misrepresentation of teaching practices to "look good" as well as unintentional misreporting due to a teacher's perception that he or she is correctly implementing a practice when in fact it is not being implemented with fidelity. Potential biases may lead to both overreporting and underreporting of practices, making the data difficult to interpret. Although this phenomenon has been widely researched in the psychology literature, more research is needed to determine the extent of its effect in the context of education and teaching. Some of the inconsistency caused by socially desirable responding may be controlled by ensuring the confidentiality and anonymity of teacher responses, gathering data longitudinally rather than just at one point in time, and gathering data from more than one source. However, these measures are not likely to eliminate all bias (Spector, 1994).

Several additional concerns warrant attention when selecting, designing, or administering self-report measures. An issue raised by several studies is the importance of ensuring consistent interpretations of terminology and a shared understanding of what the measures entail (Ball & Rowan, 2004; Blank et al., 2001; Mullens, 1995). This may require training of both teachers and outside raters (if applicable) on the survey or log measure in order to elicit the intended information. In addition, consideration should be taken to determine how broad or how detailed a survey needs to be to inform its desired purpose. Mullens (1995) notes, "Because the number of questions and the respondent burden by necessity must be limited, …indepth questions often preempt items representing a broader range of inquiry and may result in specific and often detailed information about a relatively narrow range of interest" (p. 18). Conversely, gathering information on a wider range of topics or practices may result in an insufficient amount of detail. Blank et al. (2001) also make the point that selecting a random and/or representative sample and ensuring high response rates are important considerations for obtaining valid self-report measures. Their study indicates that response rates were highest when teachers were given in-house time and support to complete the measures. In addition, teachers were more likely to complete measures when they received something of personal value from the process. Blank et al. (2004), therefore, recommend providing teachers with results that may inform their practice and assuring teachers that responses are confidential and will not be used in any way for accountability purposes.

Student Ratings

Description

It can be argued that student opinions of a teacher are an important consideration in any teacher evaluation system because students have the most contact with teachers and are the direct consumers of a teacher's services. Given their extensive experience with teachers, it seems that

valuable information can be obtained through student evaluations of teachers in the form of surveys or rating scales. However, student ratings of teachers are sometimes not considered a valid source of information because of potential biases that may affect their ratings and lack of knowledge about the full context of teaching. For example, studies have investigated whether student ratings are influenced by student age or academic level, expected or actual grades, and level of course challenge (e.g., Worrell & Kuterbach, 2001). As with teacher self-report measures, the reliability and validity of student ratings depend to some extent on the instrument used, how it is developed, how it is administered, and the level of detail it attempts to measure. The following example studies investigate the validity of student ratings for evaluating teachers.

Examples

Peterson, Wahlquist, and Bone (2000) examined whether student ratings could provide reliable and valid information to teacher evaluation. An item analysis of 9,765 student surveys, which varied by grade level (primary, elementary, and secondary), showed that students responded reliably and validly when rating their classroom teachers, though scores tended to be skewed toward high satisfaction. The study also revealed that students of different age groups may focus on different aspects of teaching. Findings showed that younger students were more concerned with teacher-student relationship (e.g., "teacher shows caring and respect"), whereas older students placed more weight on student learning. The study also reported that teachers were favorable toward having student ratings as one part of their larger evaluation system, attesting to the face validity of student ratings.

There is also evidence that student ratings can be valid predictors of student achievement. A study of schools in Cyprus by Kyriakides (2005) included a student survey of teacher practices in which the rating scales relating to teacher-student relationship and the degree of cooperation between teacher and students were highly correlated with achievement gains for mathematics and Greek language as well as with affective outcomes of schooling. In a study that compared principal ratings, student ratings, and teacher self-ratings to measures of student achievement on criterion-referenced tests in mathematics and reading, Wilkerson et al. (2000) found that student ratings were more highly correlated with student achievement than the other ratings and were the best predictor of student achievement across all subjects. These studies provide convincing evidence that student ratings of teaching are worth considering for inclusion in teacher evaluation systems.

Strengths and Cautions

There are several persuasive arguments for considering student ratings of teachers as part of the teacher evaluation process. In an empirical literature review on using public secondary school students' ratings to evaluate teachers, Follman (1992) notes that students are the most direct clients of teachers and, thus, have a broader and deeper experience with teachers than other potential evaluators, including principals, administrators, peers, or parents. A teacher's first responsibility is to his or her students, and students are in turn the most frequent source of feedback on a teacher's performance. Follman (1992, 1995) goes on to conclude that although validity concerns, such as rating leniency and halo effects (i.e., when an opinion on one trait or aspect of teaching influences all other ratings in the same direction) may affect student

evaluations of teaching, they do not seem to affect students more so than adult raters. Secondary students were shown to be capable of providing reliable ratings, validly reporting classroom events and teacher interactions, and judging whether or not a teacher is "meritorious."

In a study showing that high-achieving secondary school students could rate teaching behaviors as reliably and validly as college students, Worrell and Kuterbach (2001) note that student ratings are cost-efficient and time-efficient, can be collected anonymously, and can be used to track changes over time. They also require minimal training, though employing a well-designed rating instrument that includes detailed items measuring meaningful teacher behaviors would be important in maintaining the validity of the results.

However, researchers caution that student ratings should not be stand-alone evaluation measures because students are not usually qualified to rate teachers on curriculum, classroom management, content knowledge, collegiality, or other areas associated with effective teaching (Follman, 1992; Worrell & Kuterbach, 2001). Overall, the reviewed studies recommend that student ratings be included as part of the teacher evaluation process but not as the primary or sole evaluation criterion.

Value-Added Models

Description

Value-added measures provide a summary score of the "contribution of various factors toward growth in student achievement" (Goldhaber & Anthony, 2003, p. 38). Value-added models can be defined of "a collection of complex statistical techniques that use multiple years of students' test score data to estimate the effects of individual schools or teachers" (McCaffrey et al, 2003, p. xi). Although value-added models also may be used to evaluate schools for accountability purposes, this research synthesis concerns their use for evaluating teachers in terms of their effectiveness relative to other teachers.

Measuring effectiveness at the classroom level, rather than at the school level, is increasingly the focus of effectiveness research (Creemers & Reezigt, 1996). Researchers have focused on trying to determine teacher effectiveness by examining teachers' contribution to student achievement gains for many years, but a lack of valid measures and instrumentation has hampered the process. Only in the last 10–15 years have researchers had the necessary combination of sufficient computing power, extensive data on student achievement linked to individual teachers, and appropriate statistical models with which to determine effectiveness in terms of teachers' contributions to student learning. The result is a set of sophisticated statistical models that are used with linked student-teacher data to measure teachers' contributions to the student achievement growth of the students they taught in a given year.

Value-added models are promising, controversial, and increasingly common as a method of determining teacher effectiveness (when effectiveness is construed as teachers' contributions to achievement). However, it is also the method that is the least understood by most education professionals and teachers. Unlike classroom observations in which the teachers actually meet their evaluator, value-added model evaluators conduct their analyses from afar.

The models are complex; however, the underlying assumptions are straightforward: students' prior achievement on standardized tests can be used to predict their achievement in a specific subject the next year. Whether the student met, exceeded, or failed to reach the predicted score forms the basis for the teachers' effectiveness score. When most students in a particular classroom perform better than predicted on standardized achievement tests, the teacher is credited with being an effective teacher, but when most students' in a particular classroom fail to meet predicted gain scores, the teacher may be deemed less effective. In some models, students' prior achievement scores are the basis for calculations of effectiveness, whereas other models include students' gender, race, and socioeconomic background, and still others include information about teachers' experience.

Examples

Heneman, Milanowski, et al. (2006) conducted a multiyear mixed-methods study investigating the validity of teacher evaluation systems in four sites throughout the country. The instruments they examined were modifications of Danielson's (1996) *Framework for Teaching* and included planning and preparation, the classroom environment, instruction, and professional responsibilities. They used a value-added model in which achievement was estimated based on prior achievement and other student characteristics and found positive relationships between teacher evaluation scores and student achievement gains, although there was substantial variability across sites (and within sites). Although the study focused on the evaluation instruments, there was a fairly high correlation in two sites between what the teachers were observed to be doing in their classrooms and the achievement gains of their students. The authors theorized that the higher correlation was likely due to using multiple evaluators and, in Cincinnati, highly trained evaluators. At the sites with lower correlation, there was a single evaluator with less training conducting the evaluations.

Heneman, Milanowski, et al. (2006) focus attention on one type of research that may prove to be useful in establishing the validity of various measures of teacher effectiveness. This type of research correlates scores on various measures to draw conclusions about the information the measures can actually provide. For example, they speculated that finding links between what teachers did and student test scores was in part dependent on the performance of the classroom evaluators, not just the performance of the teachers. Although they did find some connection between teachers' performance and student test scores, the findings were not consistent across sites, suggesting that using value-added strategies instead of classroom observations as a measure of teacher effectiveness does not necessarily result in more valid assessments.

Holtzapple (2003) used Danielson's (1996) *Framework for Teaching* to compare student achievement with teachers' evaluation scores using a value-added model of predicted achievement versus actual achievement in Cincinnati. The author found a correlation between the observation scores and the value-added scores for teachers: teachers who received low ratings on the instructional domain of the teacher evaluation system had students with lower achievement, teachers with *advanced* or *distinguished* rankings on this instrument generally had students with higher-than-expected scores, and teachers rated *proficient* had students with average gains.

Interestingly, one of the sites investigated by Heneman, Milanowski, et al. (2006) was Cincinnati, and it was one of the sites that had higher correlations between the observations and the value-added scores. Cincinnati had highly trained raters conducting evaluations, which may explain the correlation. If, in fact, observable teacher practices lead to improved student learning, then there certainly should be a correlation between these two measures.

A similar study by Kimball et al. (2004) examined the relationship between teacher evaluation scores and student achievement in nine grade-test combinations in one county. Using an adaptation of Danielson's (1996) *Framework for Teaching,* this study estimated teacher effects on student achievement and determined that teacher practices contributed slightly to student achievement. However, only two of the correlations were statistically significant. This finding suggests that there is still much to learn about what value-added models are actually measuring because the research is not providing strong, consistent correlations between what teachers do in their classrooms and value-added scores.

Other researchers have calculated value-added scores for teachers and then tried to correlate them with other explanatory information. For example, Aaronson, Barrow, and Sander (2007) conducted a study using Chicago public high school data, focusing on mathematics. They calculated value-added scores for teachers and then attempted to correlate these scores with teacher characteristics including age, experience, degree level, certification, and undergraduate major. They found that almost none of the variance in teacher effectiveness—except having an undergraduate major in mathematics or science—was accounted for by these characteristics. The authors concluded that the differences in teachers were not to be found among the teacher characteristics for which they had data. This study demonstrates an unfortunate fact about value-added scores—they reveal nothing about *why* teachers vary in their effectiveness as measured by student achievement score gains. Thus, it is impossible to either predict which teachers will be most effective or help less effective teachers improve.

Rivkin et al. (2005) attempted to correlate observable teacher characteristics, such as education and experience and unobservable components to student achievement gains in Texas. They determined that observable teacher characteristics have small but significant effects on student achievement gains but found that the majority of teacher effectiveness cannot be explained by these observable characteristics. In other words, they demonstrated that teachers vary in their contribution to students' achievement score gains, but they could not explain what caused the variation. Again, this study points out a key problem with value-added measures—they do not enhance understanding of what effective teachers do that makes them effective.

Another study focused on whether teachers fostered student creativity in their classrooms and used observation scores as predictors of student achievement gains (Schacter, Thum, & Zifkin, 2006). After multiple classroom observations, the researchers found that most teachers did not employ teaching strategies that encouraged students' creativity, but when they did, the result was improved student achievement. This study illustrates an important point about using value-added models: High-quality observational data, when combined with a sound value-added model, may provide useful information about differences in teaching that could lead to strategies for improving student outcomes. In this instance, if the teacher behaviors that promoted student

creativity could be taught to other teachers, better student achievement might result. On the other hand, without the observational data, the authors would know only that students of some teachers had better achievement gains—but they would not know what practices were responsible for those differences. Clearly, value-added models have great potential for improving instruction when combined with observational data, though there are still questions to be answered. Chief among them is how to sort out the impact of one particular variable—teaching for creativity, for instance—from all of the other interactions between teachers and students that lead to learning.

Value-added models also are being used for research projects examining teacher preparation programs, such as the Carnegie-funded Teachers for a New Era (Sanders & Rivers, 2006) and Louisiana State University's value-added assessment of teacher preparation (Noell, Porter, & Patt, 2007). The goal of these studies is to better understand the relationship between what teachers learn in preparation programs and their students' achievement gains. Unfortunately, the inability to get appropriate longitudinally linked student-teacher data has hampered such efforts.

Brief summaries of other studies appear in the appendixes. There is little validity evidence linking value-added scores to teacher characteristics or practices—or even to school characteristics or practices. Teachers vary greatly—even within schools—in their effectiveness as measured by standardized test scores, but that variation has not been consistently and strongly linked to what teachers do in their classrooms. This suggests that either classroom observation instruments are not sensitive enough to capture the differences that matter in terms of student achievement or that other things are being measured that have not yet been conceptualized. So, although it is possible to say that students in one classroom learned more than students in another, it is not possible to say with any certainty why that occurred. Thus, value-added models are limited in their usefulness because the information gleaned from them is essentially a "black box"—the classroom context and teacher characteristics, qualifications, and practices that produced the value-added scores are unknown. This speaks to the importance of having additional components of a useful system of evaluating teacher effectiveness.

Strengths and Cautions

Value-added models are a relatively new way to measure teacher effectiveness, and there are researchers who support their use (e.g., Hershberg et al., 2004; Sanders, 2000). These researchers argue that value-added models provide an objective means of determining which teachers are successful at improving student learning as measured by gains on standardized tests. It is possible for teachers evaluated with a classroom observation instrument to receive a high score but still have students with average or below-average achievement growth. In addition, observation instruments can be used to evaluate teachers on their use of teaching practices that reflect experts' beliefs about good teaching, but there is a dearth of empirical evidence that specific teaching practices improve student learning (see Goe, 2007, for a synthesis of this research). This mismatch between what teachers do in their classrooms and student achievement gains may be due in part to the difficulty of measuring differences in teaching practices with standardized achievement outcomes (see Valli et al., 2004, for a discussion of these difficulties). Because value-added measures focus only on actual student gains on standardized tests, the extent to which teachers' practices reflect an instructional ideal is not relevant. Under this model,

teacher effectiveness is based on confidence that student test scores are valid and reliable indicators of student learning.

Value-added results may be able to help identify exemplary teachers. Across schools and even within schools, there are considerable differences among teachers in terms of their contributions to student learning (Rivkin et al., 2005; Rockoff, 2004). New or struggling teachers may benefit by observing highly effective teachers, but these outstanding teachers are often identified through their reputation. Value-added scores provide a means to identify highly effective teachers whose practices contribute the most to student learning gains. Establishing these teachers' classrooms as "learning labs" for colleagues and researchers may provide valuable information about what practices and processes contribute to student achievement gains. It would be especially useful to identify—and learn from—teachers who are successfully teaching students who are at-risk for poor educational outcomes.

Despite these potentially positive uses for value-added models, some researchers express reservations and describe serious concerns about their use for assessing teacher effectiveness (e.g., Bracey, 2004; Braun, 2005b; Kupermintz, 2003; McCaffrey et al., 2003; Thum, 2003). In his critique of value-added models, Bracey (2004) makes an interesting point: "V[alue added assessment] is not a theory of what makes a good teacher in all the complexity that that might require. It was developed as an atheoretical method, a technology" (p. 333). Bracey highlights a key issue of using value-added methods as a means of evaluating teacher effectiveness—that good teaching is complex, and the "technology" of value-added models examines what appears to be the results of that complex process, without regard to the causes.

Heubert and Hauser (1999), in their National Research Council report on high-stakes testing, recommended that, "accountability for educational outcomes should be a shared responsibility of states, school districts, public officials, educators, parents, and students" (p. 3). Using value-added models as the primary means of evaluating teacher effectiveness is not recommended because it holds teachers solely accountable for achievement, rather than including others who also contribute to student outcomes. Using a single score for a teacher as a measure of his or her effectiveness suggests that all, or nearly all, of the student learning in a particular subject or classroom in a given year was the product of a single teacher's efforts.

It is not just the use of value-added models that is subject to cautions from researchers. Berliner (1976) discussed the "obstructions to the study of teacher effectiveness," identifying the lack of "replicable findings relating teaching behavior to student achievement in natural classroom settings" as a key issue and noted that "instrumentation problems connected with the independent and dependent variables commonly used in research on teacher effectiveness" (p. 5) made data collection and analysis problematic. More than 30 years later, the same "obstructions" hamper the work of evaluating teacher effectiveness, particularly using student achievement to measure teacher effectiveness.

In fact, criticisms of using test scores to measure teacher effectiveness are not new. Shavelson et al. (1986) critique the process-product research that was popular in the 1970s in which researchers studied the link between teacher behaviors and student outcomes. Their appraisal of the process-product research focuses on the following four factors:

- Perfect alignment of local curriculum with the standardized test is assumed, when such alignment seldom exists, resulting in teachers being judged by their adherence to the *test's* curriculum.

- Standardized tests are strictly summative, but summary scores are not adequate reflections of improvements in students' cognition; thus, important information about students' capacity for understanding is not tested.

- Students' performance on the test is equated with their knowledge of the subject, even though the tests may be inaccurate measures of that knowledge, due to motivation, test-taking strategies, and attitudes toward testing—all of which are "extra-knowledge" influences that may affect test scores.

- Aggregating test scores across all students in a classroom may mask teachers' contributions to student learning by ignoring differential learning among students that actually reflects teachers' abilities to target appropriate instruction based on individual needs.

Shavelson et al. (1986) argue for measuring teacher effectiveness in ways that "attend to the organization of instruction in classrooms and differences in students' reactions to it" (p. 57).

The concerns about what value-added models can and cannot measure in terms of teacher effectiveness have not prevented the growth of value-added models as a seemingly objective measure of teacher effectiveness. Many states—including North Carolina, Pennsylvania, Ohio, Tennessee, Louisiana, and Florida—now use some type of value-added modeling, though they do not all use the results as a means of ranking teachers. However, an increasing number of states and school districts are exploring the use of value-added models either instead of or as a component of their previous systems of evaluating teacher effectiveness. Given this increased use of value-added models in this way, it is important to consider whether they are valid measures of teacher effectiveness.

McCaffrey et al. (2003) have argued that incomplete data and confounding influences that impact student scores that may not be included in the models (e.g., school effects) present major challenges to using value-added models for determining teacher effectiveness. In fact, Braun (2005a) has stated that what are typically called "teacher effects" are more accurately termed "classroom effects." This distinction is made because student learning is impacted by many variables in classrooms besides the teacher. It is not possible to sort out what part of a students' growth (or lack of growth) is solely attributable to the teacher's efforts. Thus, it is possible to see that students in one classroom had greater gains in achievement; however, the statistical models reveal nothing about why this is so, nor how much of the difference in student gains was due to effective teaching rather than other variables.

Another issue that has been raised by researchers is the impact of value-added measures of nonrandom assignment of students to teachers. Students are assigned to teachers by a number of methods. Different schools use different strategies, but the result is that the students in a given classroom were likely assigned to that classroom for a reason. If all students were randomly assigned to classrooms, there would be much more confidence in the resulting scores from the

use of value-added models. Several researchers have conducted studies that examine the impact of nonrandom assignment on value-added scores and concluded that there are no currently used models that adequately deal with the problem of nonrandom assignment (e.g., Rivkin & Ishii, 2008; Rothstein, 2008a, 2008b).

Finally, the validity of using value-added models for measuring teacher effectiveness is dependent in part on whether the statistical models are correctly specified and whether the inferences drawn are appropriate and defensible. The causative elements are not usually included in the modeling. Teachers teach, but *what* and *how* they teach are not part of the statistical model. So even though it has been determined that teachers differ in effectiveness in terms of producing student learning gains, ways to replicate those differences are not apparent. Even if teachers could be cloned, the teaching context (students, curriculum, resources, parental support, school leadership, etc.) would vary. Teachers may be differentially effective (i.e., a teacher who is successful in one context may be less successful in another).

Toward a Comprehensive View of Teacher Effectiveness

In many states, teacher effectiveness is assessed by focusing on results from a single measure, typically classroom observations and less commonly, teachers' contributions to student achievement growth (value-added models are one mechanism for examining this growth). Revisiting the five-point definition of teacher effectiveness, it is clear that using one or even both of these methods of measuring teacher effectiveness fails to indicate the many important ways in which teachers contribute to the success and well-being of their students, classrooms, and schools. Thus, creating a comprehensive score for teachers that includes multiple measures is one possible way to capture information that is not included in most classroom observation protocols or in scores developed using value-added models.

What types of measures might be included in this comprehensive measure? Here are some options for collecting data (from New Mexico's teacher performance evaluation guidelines): "review of videotape (of lesson); written documentation of activities; locally developed survey of staff, students, and/or parents; review of student work and performance; review of the teacher's contribution to the school's vision, mission, and outcomes; portfolios; information gained through peer observation and/or peer coaching; anecdotal records; reflective journals; self-evaluations; instructional artifacts; other formats" (New Mexico 3-Tier Licensure Implementation Teacher Training Work Group, 2005, p. 9). Unfortunately, there is little empirical evidence of the validity of these various methods for measuring teacher effectiveness, and in many cases, there are no standardized instruments for data collection. Instead, the collection of data—and decisions about what is important to collect—is left up to local decision makers.

Considering Teaching Contexts

Deciding how teacher effectiveness should be measured is not necessarily the sole purview of policymakers, researchers, and bureaucrats. Given that teaching contexts vary widely, it is essential that local input is considered when decisions are made about what to prioritize in a composite measure of teacher effectiveness. For example, a district with a high percentage of English language learners may want to consider teachers' ability to communicate effectively with these students and their parents as part of their composite measure of teacher effectiveness. Similarly, an urban school that has a high proportion of student dropouts may want to include a measure of teachers' documented efforts to assist at-risk students as part of their composite measure of teacher effectiveness. And a school in which teacher collegiality has been lacking might want to consider evidence of ways in which teachers initiate, lead, or support efforts to work together in professional learning communities.

Given that instruments and protocols for measuring teachers' leadership activities or contributions to improvement in school climate have yet to be developed in some cases, and standardized in most cases, it is not possible to make recommendations about what a state or local education agency should include in the creation of a valid composite measure of teacher effectiveness. Rather, it is recommended that the definition of teacher effectiveness be broadened, that it be inclusive of state and local priorities, and that it consider teaching contexts. Obviously, some schools have little or no problem with student attendance or dropouts, whereas other schools may lose days of students' learning time or lose students altogether. In some

schools, then, a measure of ways in which teachers have worked toward improving attendance or preventing students from dropping out would be a low priority, whereas such a measure would be a high priority in other schools.

Another consideration is that teaching contexts differ greatly across subjects and grades, and some types of measures may be more suitable for certain types of contexts. Campbell et al. (2003) critique teacher effectiveness models that are applied equally to all school levels and contexts, without regard to what may distinguish effectiveness in a particular subject, grade, or context. They argue for incorporating five dimensions of differential teacher effectiveness: "differences in activity, differences in subjects and/or components of subjects, differences in pupils' background factors, differences in pupils' personal characteristics, differences in cultural and organisational context" (p. 354).

Most classroom observation protocols, including Praxis III and Charlotte Danielson's (1996) *Framework*, are intended for use in all classrooms without regard to context. The CLASS instrument, however, has a Grades PK–3 version that has been extensively tested (La Paro, Pianta, & Stuhlman, 2004) as well as a more recently developed middle- and secondary version that is currently being piloted (Pianta, Hamre, Haynes, Mintz, & La Paro, 2007). These different versions of CLASS take into account the differences in teaching contexts at those levels. However, it may be possible to use a single instrument to evaluate teachers in different subjects, grade levels, and school contexts. The differences would then have to be accounted for in the scoring rather than in observation.

In their choice of teaching preparation programs, teachers select a grade level and subject in which they feel they have the most to offer their students. In their choice of schools, teachers select a context in which they feel they are likely to be successful. Yet many evaluation instruments do not acknowledge that teachers may be differentially successful depending on the context. What does this mean in terms of teacher effectiveness? First, teachers are not interchangeable—a teacher that performs well in one classroom may feel challenged in another classroom. Thus, an evaluation of teacher effectiveness should be specific to a context, subject, and grade level, and teachers should be compared with or ranked against teachers who are in similar contexts, subjects, and grade levels. In addition, evaluating a secondary science teacher's effectiveness on the same scale as that of a kindergarten teacher's effectiveness may be problematic, particularly if there is a need to identify exceptional teachers in specific contexts, grades, or subjects. This need might arise from a number of situations, including identifying a suitable mentor for a novice middle-school ELA teacher, rewarding exceptional teaching at the elementary level, recruiting teachers who have proven to be especially able to work with at-risk students for a special program within a school, or even offering an incentive for transferring to a hard-to-staff school. Lastly, taking into account teachers' evaluations when making hiring or transfer decisions might ensure a better match to open positions. A teacher's record of effectiveness in a specific setting may be a factor worth considering.

Using Teacher Effectiveness Results to Improve Instruction

There are many different purposes for evaluating teacher effectiveness; a key reason is to identify weaknesses in instruction and develop ways to address them. For this reason, one goal of evaluating teaching effectiveness should be to collect information that will be useful in designing appropriate strategies to improve instruction. Approaches to improving instruction may involve professional development, individualized work with a curriculum specialist, college coursework, and study teams within or across schools. Smylie and Wenzel (2006), citing a number of successes among school districts around the country, recommend a "human resources management" approach to improving instruction, wherein vertical and horizontal alignment of practices enable school leaders to carry out instructional objectives. They reported on three Chicago elementary schools that coordinated and aligned human resources to improve practices, including "teacher recruitment and induction, professional development activities, communication of expectations for teacher performance, specification of classroom teaching strategies, provision of encouragement and incentives, principal supervision and evaluation, and removal of poorly performing teachers" (p. 24).

Other sites may choose a more individualized approach to improving instruction, allowing teachers to plan their own professional growth. Denver's Professional Compensation System (ProComp) is an example of a district that has created a sophisticated system that permits considerable flexibility for teachers to decide how they will improve instruction (for additional information, see the ProComp website at www.denverprocomp.org). In collaboration with principals and supervisors, teachers can create a plan for their professional development, including taking courses (with tuition reimbursement) that will address gaps in their knowledge. Teachers and their supervisors can use evaluation results (from classroom observations and student achievement gains) to help them determine areas that need to be addressed.

Although there are many possible approaches besides those mentioned, the point is that evaluating teacher effectiveness should ultimately lead to improved instruction. In addition, under the broad definition of teacher effectiveness presented in this synthesis, evaluations also can be used to identify other areas in which teachers are performing well or they may need additional support. For example, if a district's priority is decreasing referrals to special education by identifying and providing assistance to at-risk students, it may be necessary to create opportunities for teachers to collaborate with colleagues and other education professionals during the school day.

A Final Note About Validity

When designing systems for evaluating teacher effectiveness and using the results of such evaluation, it is important to keep in mind that ways of measuring teacher effectiveness—such as classroom observation protocols or value-added models—are not valid in and of themselves for determining teacher effectiveness. Rather, their validity lies in their ability—when used correctly—to accurately and reliably measure what they were intended to measure. For classroom observation instruments, validity lies in the instrument's ability to measure how well a teacher exemplifies standards of practice that have been deemed important for that grade level, subject, and teaching context by some group of experts. For value-added measures, validity lies

in how well the model accurately captures an individual teacher's contribution to student achievement growth in a particular subject area.

At this juncture, researchers still have a long way to go toward clearly establishing the validity of various instruments for the purpose of measuring teacher effectiveness. There have been many research studies published to establish the validity of various measures of teacher effectiveness (e.g., examining how a score from an observation instrument correlates with a value-added score); however, validity cannot be determined by correlating results from measures based on two different constructs. Rather, validity must be determined by how well a given teacher's performance matches the construct—whether that means keeping at-risk students in school, contributing to a positive classroom environment, or having a high value-added score. Thus, the crucial step in getting valid information is deciding what is important and then finding (perhaps creating) a measure that will yield concrete evidence about teachers' performance on what is important. In a broad definition of teacher effectiveness, such as the one suggested, there is no single measure that will provide valid information on all the ways teachers contribute to student learning and growth and to their schools. Multiple measures—each designed to measure different aspects of teacher effectiveness—must be employed.

Policy Recommendations and Implications

The following set of recommendations is designed to provide guidance to entities that are considering how best to measure teacher effectiveness:

- Resist pressures to reduce the definition of teacher effectiveness to a single score obtained with an observation instrument or through using a value-added model. Although it may be convenient to adopt a single measure of teacher effectiveness, there is no *single* measure that captures everything important that a teacher contributes to educational, social, and behavioral growth of students, not to mention ways teachers impact classrooms, colleagues, schools, and communities.

- Consider the purpose for the evaluation of teacher effectiveness before deciding on the appropriate measure to employ. Scores from a value-added model may provide information about a teacher's contribution to student learning, but it would be less helpful in providing teachers with guidance on how to improve their performance.

- In considering the validity of various ways of measuring teacher effectiveness, keep in mind that the validity does not lie solely with the quality of the instrument or model but also with how well the instrument measures the construct and how the instrument is used in practice. Even a good classroom observation instrument in the hands of untrained evaluators may result in vastly different scores for similar teacher practices. And using a value-added model when large amounts of student data are missing may yield scores that fail to reflect the teacher's actual contribution to student learning.

- Seek other measures, or create appropriate measures, to capture important information about teachers' contributions that go beyond student achievement score gains. This may mean developing a measure that captures evidence of an individual teacher's leadership activities within the school, his or her collaboration with other teachers to strategize ways to help students who are at risk for failure, or participation in a study group to align the curriculum with state standards.

- Include education stakeholders in decisions about what is important to measure. Although a state legislature or task force may ultimately decide upon how teacher effectiveness will be measured, listening to the voices of teachers, principals, curriculum specialists, union representatives, parents, and students will help assure greater acceptance of the measurement system. Ultimately, this also will contribute to greater validity; the validity of a measure can be threatened by noncompliance or active resistance to the measure.

- Keep in mind that valid measurement may be costly. Ensuring that data is complete and accurate and that raters are trained and calibrated is essential in order to ensure the validity of the scores of the most commonly used measures of teacher effectiveness. Developing and validating new measures based on local priorities also will require adequate funding.