

Date: October 5, 2009

To: Brian D. Austin, Program Director and Founder, The Animation Project

From: Kathyne B. Brewer, Rachel Wetts, Allon Yaroni, Vera Institute of Justice

Subj: Evaluability Assessment

The following memo outlines the findings of an evaluability assessment of The Animation Project (TAP) conducted by the Vera Institute of Justice (Vera). The concept of evaluability assessments evolved during the 1960s and 1970s, when researchers noted that conducting an evaluation prematurely might either generate flawed conclusions about program success and failure and/or fail to produce relevant information for program improvement. Troubled by this finding, the Urban Institute developed a systematic process for determining whether a program is ready for evaluation.¹ Evaluability assessments are frequently used today as a precursor to full-scale evaluations. The purpose of these assessments is to help researchers and program managers explore the feasibility and potential usefulness of an evaluation of the program.

The evaluability assessment permits researchers to determine whether a program has implemented the basic foundations necessary for an evaluation. Evaluability assessment seeks to answer the following questions: (1) Does the program have a clearly articulated design and/or logic model with plausible connections between client needs, program activities (outputs), and intended goals (outcomes)? (2) Does the program serve the population it is intended for? (3) Does the program possess the resources (inputs) for implementation? (4) How is the program being implemented? With this information, researchers and program stakeholders are able to determine whether a program or policy is ready for evaluation, if features of the program require refinement before the program can be evaluated, or whether a program is completely ineffective and should be stopped altogether.

The process of developing a refined logic model for TAP (see Appendix A) as well as reviewing best practices in the field has provided valuable information about the program—including objectives, key program components and their implementation, and theoretical links

¹ Trevisan, M. & Yi Min, H. (2003). Evaluability Assessment: A primer. *Practical Assessment, Research & Evaluation*, 8(20), Retrieved July 7, 2009 from <http://PAREonline.net/getvn.asp?v=8&n=20>.

Van Voorhis, P. & Brown, K. (1996). *Evaluability assessment: A tool for program development in corrections*. Unpublished monograph, National Institute of Corrections, Washington, DC.

between the components and desired outcomes—that form the foundation of the evaluability assessment. Findings concerning program design and implementation that are relevant to the program’s evaluability are as follows:

1. TAP has developed a coherent program design. The program’s client population, key activities, and desired outcomes have been articulated in a formal logic model (see Appendix A). The model’s posited connections between program resources, activities and expected outcomes are supported by theoretical and, where possible, empirical literature. Experts in fields related to TAP’s work who were interviewed have pointed out that the program’s goals are realistic to its client population and to its methods. Consensus about the program’s goals and priorities, as articulated in the logic model, has been reached between members of program staff.

2. TAP serves the population for whom it was designed. TAP aims to serve at-risk youth 12 to 18 years of age. By partnering with The Family Center and CASES, TAP accesses and serves communities of youth with risk factors including juvenile justice involvement and family health and bereavement issues. Young people served at MS4 were referred to the program through the school’s art therapist, who hand-picked youth who were at risk and for whom “other modalities haven’t worked.”

3. TAP possesses the staff and resources called for in its program design. TAP possesses connections to community service providers who provide client referrals and program space. Technical resources (i.e., computers and animation software) have been purchased to meet current operational needs. TAP also possesses personal and business connections to animation professionals to provide access to internship opportunities for youth. Each member of TAP’s staff has appropriate academic training and professional experience in their area of expertise.

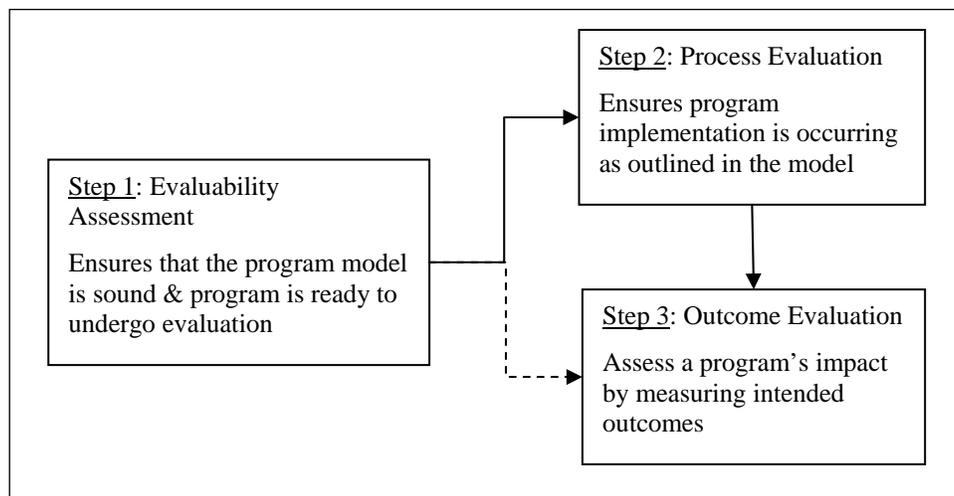
4. All activities in the program design are being implemented, though with significant variation across program sites. Through interviews with TAP program staff and observations of group sessions, we have confirmed that TAP is implementing all activities outlined in its logic model. However, the extent to which these activities are implemented varies by site. For example, technical animation skills training is emphasized at the MS4 site, but is minimal at The Family Center and CASES sites, while working in groups is not emphasized at the MS4 site as it is at the other two sites. Such variation is at least partially a result of TAP’s decision to use this year to pilot various approaches to service delivery as it tests its model. TAP program staff have articulated in meetings and interviews that program implementation will be more uniform in the upcoming year.

Based on this information, we believe that both process and outcome evaluations of TAP would be appropriate. The following sections detail the main steps that TAP would need to take in order to accomplish these evaluations, including an examination of data collection needs, the

tools required to meet these needs (e.g., intake and data collection forms and instruments), and potential design strategies for conducting the evaluation.

We begin by presenting the steps TAP needs to take to accomplish a process evaluation. While assessing program impact is the ultimate goal of a series of evaluations, outcome evaluations are based on the assumption that a program is operating as designed. As a result, we strongly recommend conducting a process evaluation either prior or concurrent to an outcome evaluation in order to ensure that the results of an outcome evaluation are accurate. If the program's activities are not being implemented as designed, an outcome evaluation may report no results or even negative results, even though the program's model is efficacious. Therefore, a process evaluation may be used or implemented without an outcome evaluation, but an outcome evaluation without a process evaluation may lead to results that are invalid, or even injurious to a program's continuation. In TAP's case, a process evaluation will be particularly helpful, since TAP's program model is being implemented differently across sites.

Figure 1: Recommended Steps



Process Evaluation

Based on the evaluability assessment, TAP is ready for a process evaluation. Process evaluation uses empirical data to evaluate the delivery of a program, examining whether the program is being implemented as designed (i.e., according to the logic model) and identifying any gaps that exist between program design and practice. This is an important step to take either prior to or concurrent with an outcome evaluation as the extent to which the logic model is being implemented strongly affects TAP's ability to measure the relationship between the program and the outcomes.

Study Design

We recommend a three-step process evaluation which is flexible in design, and where each phase of the evaluation would depend upon the results of the prior phase. Each successive phase is not required, and only the first phase (the fidelity assessment) needs to be completed to ensure that an outcomes evaluation will produce results that may meaningfully assess whether the program model is efficacious. However, additional steps taken in the process evaluation will lead the outcomes evaluation to have more explanatory power, and will provide TAP staff with further insight into how their program might be improved. Figure 2 outlines the research questions of each of the three phases, the type of findings they are designed to produce, and the action steps appropriate for those findings.

Phase 1 - fidelity assessment. The first and fundamental research question that a process evaluation seeks to answer is whether the program is being implemented in accordance with the logic model. This step in the evaluation can constitute the entire process evaluation, or may yield results that prompt further investigation. For example, if a fidelity assessment finds that a certain program activity is not being implemented, the next step in a process evaluation would most likely be attempting to determine the reason for this lack of fidelity.

Phase 2 - variations in implementation. This assessment seeks to uncover variations in how TAP is implemented across its three primary sites (CASES, MS4, and The Family Center). Based on our observations and the findings of the current evaluability assessment, TAP is currently being implemented differently at each of these sites in regard to relative emphasis placed on different program activities. Therefore, an outcomes evaluation of TAP may produce different results at each site. It may reveal large impact on outcome variables at one site, but relatively small or even negative results at another site. The proposed additional research phase would shed light on whether such differences in outcomes were a factor of the population served, a factor of different relative emphases placed on different program activities at each site, or a combination of the two. These findings may be useful to TAP in identifying which population of youth benefits most from its program and which activities are most strongly linked to desired outcomes.

Phase 3 – qualitative analysis of implementation discrepancies. This phase is contingent upon whether differences in implementation are discovered in Phase 1 or Phase 2 of the process evaluation. Further exploration of the underlying reasons for discrepancies through qualitative interviews with TAP staff and stakeholders may prove useful to determining how such discrepancies might be resolved, or if it is even possible to resolve them, given current resources. Conversely, it might be determined that discrepancies are not weaknesses, but rather are appropriate given differences in client population, group duration, or other such factors over which TAP has limited control. In this case, the program logic model will likely need to be modified to reflect such contingencies. The findings of such a study would assist TAP in

evaluating whether any differences between the program model and its implementation at various sites can and should be remedied, and, if so, how that may best be accomplished at each site.

Figure 2: Outline of three-step process evaluation

<p>Phase 1: Fidelity Assessment: <i>Are all of the essential Output components of the logic model being implemented at all program sites?</i></p>	
<p><i>If all components are being implemented:</i></p> <p><u>Conclusions:</u></p> <ol style="list-style-type: none"> 1) Program is being implemented as designed. 2) Results of outcomes evaluation can serve as evidence of efficacy/ inefficacy of program model. <p><u>Actions:</u></p> <ol style="list-style-type: none"> 1) Proceed with outcomes evaluation. 2) Attempt to uncover variations in implementation (proceed to Phase 2). <p><i>OR:</i> Process evaluation complete.</p>	<p><i>If any component is not being implemented:</i></p> <p><u>Conclusions:</u></p> <ol style="list-style-type: none"> 1) Program is not being implemented as designed. 2) Results of outcomes evaluation can not serve as evidence of efficacy/ inefficacy of program model. <p><u>Actions:</u></p> <ol style="list-style-type: none"> 1) Halt outcomes evaluation. 2) Attempt to uncover reasons for lack of compliance with model (proceed to Phase 3).
<p>Phase 2: Assessment of Variation in Program Activities: <i>Are program activities being implemented uniformly across sites? In other words, does the proportion of group time spent engaging in different activities vary by site?</i></p>	
<p><i>If implementation of activities is uniform across sites:</i></p> <p><u>Conclusions:</u></p> <ol style="list-style-type: none"> 1) All sites are in compliance with the logic model, and implementation is uniform. 2) Any differences in the results of the outcomes evaluation may be expected to differ primarily as a function of the populations served. <p><u>Actions:</u></p> <ol style="list-style-type: none"> 1) Process evaluation complete. 	<p><i>If implementation of activities varies across site:</i></p> <p><u>Conclusions:</u></p> <ol style="list-style-type: none"> 1) While all sites are in compliance with the logic model, implementation differs across site. 2) Results of outcomes evaluation may be expected to differ across site due both to differences in the population served, and to differences in implementation. <p><u>Actions:</u></p> <ol style="list-style-type: none"> 1) Attempt to uncover the reasons behind site-specific variations, and judge whether these variations can or should be made uniform (proceed to Phase 3). <p><i>OR:</i> Process evaluation complete.</p>
<p>Phase 3: Qualitative Analysis of Discrepancies in Implementation: <i>If discrepancies exist between how the program was designed and how it is implemented or between how it is implemented at different sites, what are the reasons underlying those discrepancies? Do stakeholders perceive discrepancies to be weaknesses of implementation, or adaptive responses to conditions?</i></p>	
<p><u>Conclusions:</u></p> <ol style="list-style-type: none"> 1) Discrepancies can and should be resolved; the analysis yields action steps to accomplish this. 2) Discrepancies are not possible to resolve at this time given current resources 3) Discrepancies are appropriate; the program logic model is modified to reflect such contingencies. 	

Measurement

In order to assess whether program model is being implemented with fidelity, and (if desired) with uniformity across sites, the process evaluation must first determine the components and activities of the program, using the Outputs section of the logic model as a guide.

The first category of the ‘outputs’ section of the logic model deals with who the program is targeted at, or the program’s ‘reach.’ It is important that a process evaluation document the characteristics of program participants in order to assure that the program is impacting the group it intends to impact (necessary to the fidelity assessment, i.e., Phase 1). This data can also often be useful in hypothesizing reasons for differences in implementation across groups that can not otherwise be explained (Phase 3). Finally, if process and outcomes evaluations are being conducted concurrently, this data may also be used in conjunction with the outcomes evaluation to ensure comparability across control and test groups, and may be helpful in determining which demographic groups benefit most from the program. Since the logic model defines the program’s reach as at-risk youth (age 12-18), we recommend collecting information on the risk factors of youth in addition to age, race/ethnicity and other socio-economic characteristics. A process evaluation may serve to improve understanding of the risk factors present for youth referred to TAP at each site, particularly the MS4 site where youth’s risk factors may not be as apparent as those at CASES or the Family Center.

The second category of the ‘outputs’ section of the logic model articulates the program’s activities. In order to determine the fidelity of program implementation, these components should be classified according to which ones are essential or core to the program being implemented with fidelity to the model, and which are non-core. For example, group completion of an animation short is likely to be a core component, while a screening of the completed animation short is likely to a non-core goal that will depend on time, resources, and willingness of the youth. Measuring whether all essential activities articulated in the logic model are being implemented at all sites is necessary to ensure program fidelity, and to ensure that results of a concurrent or future outcomes evaluation are based on a program that is implemented as designed.. Table 1 outlines our recommendations for core and non-core program activities, based on our observations, review of the literature, and conversations with program staff and experts in the field.

Table 1: Classification of Core and Non-Core Program Activities

Program Activity	Core activities	Non-core activities
Group creation of animation short	√	
Facilitators model positive social interactions	√	
Life issues processed through fictional mode and ongoing therapeutic interventions	√	
Animation short completed	√	
Animation short screened		√
Discussion of animation industry and range of career possibilities		√
Training in 3D software	√	
Internship placements		√

Data Collection and Analysis

As explained before, Phase 1 of the process evaluation focuses on whether each activity is being implemented during the evaluated program cycle. Phase 2 focuses on the extent to which these activities are being implemented across the program sites. In most cases, this means recording the amount of time that is proportionally spent on each activity. Phase 3 focuses on the underlying reasons for variations and appropriate solutions.

In order to collect information on the above variables for Phase 1 and 2 of the proposed design, we recommend using semi-structured questionnaires (one to be completed by youth, another by staff), content analysis of program documents, and structured observations. For data collection purposes, some of the program components may be divided into multiple subcomponents, so that they may be more easily measured and quantified.

In order to perform a qualitative analysis of discrepancies in implementation (Phase 3), we recommend that interviews with stakeholders (including program staff, staff of partner organizations, and participating youth) be conducted. Interview questions will depend on the nature of the discrepancy identified in Phase 1 or 2 of the process evaluation, but may include reasons underlying the discrepancy, recommended methods for resolving it, and reflections on the benefits to the program of resolving the discrepancy in comparison to the costs. Table 2 lists the main variables of interest and the data collection method recommended.

Table 2: Program Variables, Subcomponents and Data Collection

Program Component	Recommended subcomponents	Recommended method(s)
<i>Reach</i>		
At-risk youth (age 12-18)	Age	Semi-structured questionnaire, to be completed by youth
	Risk factors of youth	
	Race/ ethnicity (optional)	
<i>Activities</i>		
Group creation of animation short	Group composition (e.g., number of youth per group)	Structured observations
	Level of group participation	
Facilitators model positive social interactions	Facilitator interactions with youth	Structured observations
	Youth perceptions of facilitator friendliness/ openness	Semi-structured questionnaire, to be completed by youth
Life issues processed through fictional mode and ongoing therapeutic interventions	Themes of youth animations	Content analysis of program documents (youth animations)
	Youth perceptions of how animations relate to their lives	Semi-structured questionnaire, to be completed by youth
	Discussion of life issues during group sessions	Structured observations
Animation short completed	n/a (no subcomponents needed)	Semi-structured questionnaire, to be completed by staff
Animation short screened	n/a	Semi-structured questionnaire, to be completed by staff
Discussion of animation industry and range of career possibilities	n/a	Structured observations
Training in 3D software	n/a	Structured observations
Internship placements	n/a	Semi-structured questionnaire, to be completed by staff

Some of these data collection methods are relatively time consuming, leading to a higher level of cost. Two methods in particular—content analysis of youth animations and observations of group sessions—require a relatively lengthy data collection and/or analysis process, depending on how many animations are analyzed and how many group sessions are observed. These methods are recommended because they are directly related to the core activities of the program and are expected to produce data with high level of reliability. Less expensive, alternative measures could be used, though (as is typical of research) there is a trade-off between cost and rigor.

Content analysis of youth animations aims to identify themes and trends existing in the animations, with the goal of analyzing how these themes relate to the issues youth deal with and the challenges that place them ‘at risk.’ This corresponds to the program activity “Life issues processed through fictional mode.” We recommend that this information also be captured through youth responses to an open-ended survey question about how the animation relates to their lives, goals, challenges, etc. It would be less expensive to drop the content analysis and

measure this variable based on youth responses, and this would still likely produce meaningful results, since the youth who created the animation have the most intimate knowledge of the ways that their work relates to their lives. However, there are problems with relying solely on youth response. Some youth may not want to or be able to consciously identify all the themes in their works, and this openness and/ or ability will likely vary by age, risk factor, and other participant variables which vary by program site. It would therefore be difficult to disentangle whether differences in how often youth connected the animation to their own lives was a factor of client characteristics, or in program implementation.

The other costly method recommended is observation of group sessions using a structured observation tool. This method aims to gather information on how often a number of key program activities are being performed during group sessions. Alternatively, these activities could be measured through a structured questionnaire to be completed by program staff after every group session in which they would be asked to estimate how many minutes of the previous session had been spent performing the activities listed. This method is subject to retrospective error and bias. For example, it may be difficult for program staff to assess after the fact the number of participants who actively participated in the session, and staff are likely to be motivated to overestimate the number. We recommend that the research design of a process evaluation include at least one observation at each site, but staff questionnaires may substitute for further observations, if desired.

Ethical issues are also raised by some of the research methods recommended. Filling out questionnaires may be tedious to some youth, while others may feel uncomfortable answering certain questions. These risks can be minimized by making the surveys anonymous and voluntary, informing youth that they can skip any questions they feel uncomfortable answering, and providing some compensation to youth (usually in the form of gift cards) for their time.

We recommend that data collection occur at various points of time along any group cycle. Depending on the variable to be measured, certain information is appropriate to collect at the beginning of a group cycle, during a group cycle, directly after the last group meeting, and finally after preliminary data analysis. The following table summarizes the recommended data collection times.

Table 3: Data Collection Plan for Process Evaluation

Data collection point	Method of collection	Variables collected
Beginning of group cycle	Semi-structured questionnaire (to be completed by youth)	1) Information about participants ('Reach')
During group cycle (on-going)	Semi-structured questionnaire (to be completed by program staff)	1) Activities completed during each session [if desired as substitute for observation – see observation row below for specific variables]
	Structured observations	1) Group creation of animation short 2) Facilitators model positive social interactions 3) Discussion of life issues during group sessions 4) Discussion of animation industry and range of career possibilities
At end of group cycle	Semi-structured questionnaire (to be completed by program staff)	1) Animation short completed 2) Animation short screened 3) Internship placements
	Semi-structured questionnaire (to be completed by youth)	1) Youth perceptions of facilitator friendliness/ openness 2) Youth perceptions of how animation relates to their lives
	Content analysis of program documents (youth animations)	1) Themes of youth animations
After preliminary data analysis	Semi-structured interviews (of stakeholders)	1) Qualitative analysis of differences in implementation across program sites

Survey information may be collected in traditional (pencil and paper) format, but depending on the resources available, it might be more convenient and time-efficient to use a web-based survey device such as SurveyMonkey.² Staff and youth could directly answer survey questions using TAP's computers, which would eliminate data entry time. A computer-based method also might cause youth to feel more comfortable answering personal information or giving negative feedback, since surveys would not be connected to them in any way and it would not be necessary for them to hand the surveys to TAP staff.

Both observational instruments and survey tools need to be created for this process evaluation. As is typical of process evaluations, most of the items collected here are program-specific, so the evaluator will have to create most observational and survey instruments. However, some variables such as youth participation and youth perception of facilitators'

² See: <http://www.surveymonkey.com>

friendliness/ openness have been measured in other evaluations. Appendix B contains observational and survey tools which may serve as a guide. Portions of these tools can be used by TAP without modification.

Data from observational tools would be entered into a software program such as Excel or SPSS, and the same would have to be done for all survey tools if traditional pencil and paper format was used. Frequencies would then be generated according to the site from which the data was collected. Frequencies would reveal whether all program components were being implemented in all groups and the extent to which implementation was occurring in fidelity with the logic model. The reasons for these differences would then be investigated through semi-structured interviews with stakeholders.

We recommend that qualitative information from interviews and content analysis of youth animations be coded and analyzed using a grounded theory approach, which includes an iterative search for common and unique themes that emerge both within and across each of the three program sites. For example, in the content analysis of youth animations, scenes in the animation could be coded into one or many of broad categories representing life themes portrayed in that scene such as ‘confrontation with authority,’ ‘poverty,’ or ‘romantic relationships.’ Interviews investigating differences among groups would be coded into similarly categories, in this case representing staff perceptions of underlying reasons for these differences such as ‘client characteristics,’ ‘resources available at site,’ or ‘group duration.’ Within each of the categories sub-themes could then be identified, and connections and interlocking themes could be identified and their relationships analyzed.

Outcome Evaluation

Based on the evaluability assessment, TAP is ready for an outcome evaluation concurrent to a process evaluation. Outcome evaluations are used to assess the effect of a program on its participants. This type of evaluation will provide valuable information to TAP about whether its services are having the intended impact on youth’s lives. Additionally, by evaluating multiple sites (i.e., MS4, CASES, and The Family Center) it may offer insight as to where TAP is having the greatest impact and therefore where to focus future expansion efforts. Alternatively, TAP may wish to conserve resources by focusing the outcome evaluation on the site-type that they wish to expand.

Outcome evaluations are generally more complex and expensive than process evaluations, and should be carried out only under the following circumstances: (1) there is a clear statement about what changes are expected; (2) appropriate measures are selected for assessing the expected outcomes; and (3) a system is created to collect reliable data about these outcomes. The

first circumstance is met through the short-term, medium-term and long-term outcomes listed in TAP's logic model (see Appendix A). In the sections below, we discuss recommendations for appropriate design, measures and data collection.

Study Design

The purpose of an outcome evaluation is not only to measure changes in outcomes but also to establish the causal effect of the intervention/program. Consequently, decisions about what will be measured and when are crucial. The best evaluation designs incorporate two key features. First, they measure the variables multiple times, with the first measurement taking place before the intervention/program has started to establish a baseline to use for comparison once the program has finished. Second, in addition to collecting data on a cohort of participants, they also incorporate a control or a comparison group of similar non-participants whenever possible. The outcomes that are evaluated may be short-term, medium-term, or long-term outcomes.

Short-term Outcomes. In order to assess short-term outcomes, we recommend using a simple pre and post test design (see Figure 3). In this design, variables of interest are measured at the very beginning of a group cycle (pre-test) and then again at the very end of the cycle (post-test). These results are then compared in order to assess changes that may have occurred in the variables of interest during the program period. Although this evaluation design does not allow us to determine if the observed changes were due to the intervention/program, it does allow for demonstrating an association between the intervention/program and the observed outcomes. In other words, this design allows us to establish correlation, but not causation.

The use of a random assignment of participants to the program and to a control group will greatly strengthen the evaluation's ability to establish a causal link between the intervention/program and the observed outcomes. The pre and post test design with a control group is considered the gold standard of evaluation design as it provides the greatest amount of confidence that the intervention is causing any of the observed differences. The use of a control group would work as follows. Prior to the start of a TAP cycle, a list of eligible participants at the site would be created. Youth on the list would then be randomly assigned to join either the program (test group) or the control group. Youth who are assigned to the control group would not participate in the evaluated TAP cycle but would receive all other site services. Because the use of a control group depends upon the possibility of composing a 'wait list' of similar participants, it may not be possible to implement this aspect of the design at all program sites.

Medium-term Outcomes. In order to assess medium-term outcomes, we recommend using a longitudinal study design. This type of design follows a group of participants to evaluate changes in the variables of interest by collecting data at various points in time. As the TAP logic model anticipates positive changes in educational attainment and employability, as well as a decrease in criminal justice system involvement, a study looking at these outcomes will require a longer time

frame to determine how participants have fared in these areas. As the participants at each TAP site differ in age and circumstances, the specific time frame will likely need to vary based on the specific characteristics of the site evaluated. In general, we recommend that TAP collect data at the start of cycle, the end of cycle, and at one-, three-, and five-year intervals following participation (see Figure 4). At the CASES site, where the average age is somewhat older and outcomes concerned with recidivism may be of greater importance, TAP may want to shorten the time-frame to six-months, one-year and three-year post participation. Similar to the short-term outcomes evaluation, following a control group through the evaluation will provide the most empirically sound information on how youth are impacted by TAP.

Long-term Outcomes. Long-term outcome evaluations are logistically difficult and often too expensive to implement. Consequently, while the process of evaluating long-term outcomes is very similar to the process of examining medium-term outcomes, we do not recommend TAP undertake this process as it would be prohibitively costly. In order to establish an empirical link between TAP services and the long-term outcomes specified in the logic model, such as decreased rates of poverty, unemployment, incarceration, homelessness, and mental health problems, an evaluation would need to control for the influence of multiple other events and programs that might have occurred after participation in TAP.

TAP is likely to benefit the most from focusing on short-term outcomes – where tracking the youth is easy and outcomes are strongly linked to program activities. In addition, short-term evaluations are more likely to be conducted successfully in-house with minimal assistance from outside evaluators, thus further reducing costs. As an alternative, it may be of use to follow-up with TAP alumni, particularly those who participated in internships, at some point beyond the final data collection point to speak with them about their thoughts on how participation in TAP has influenced their lives and careers.

Figure 3: Pretest-Posttest Control Group Design

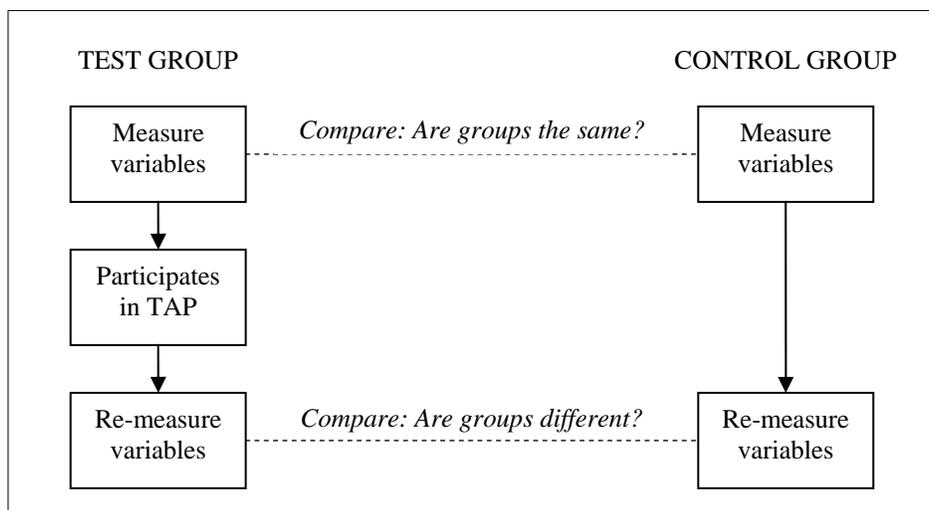
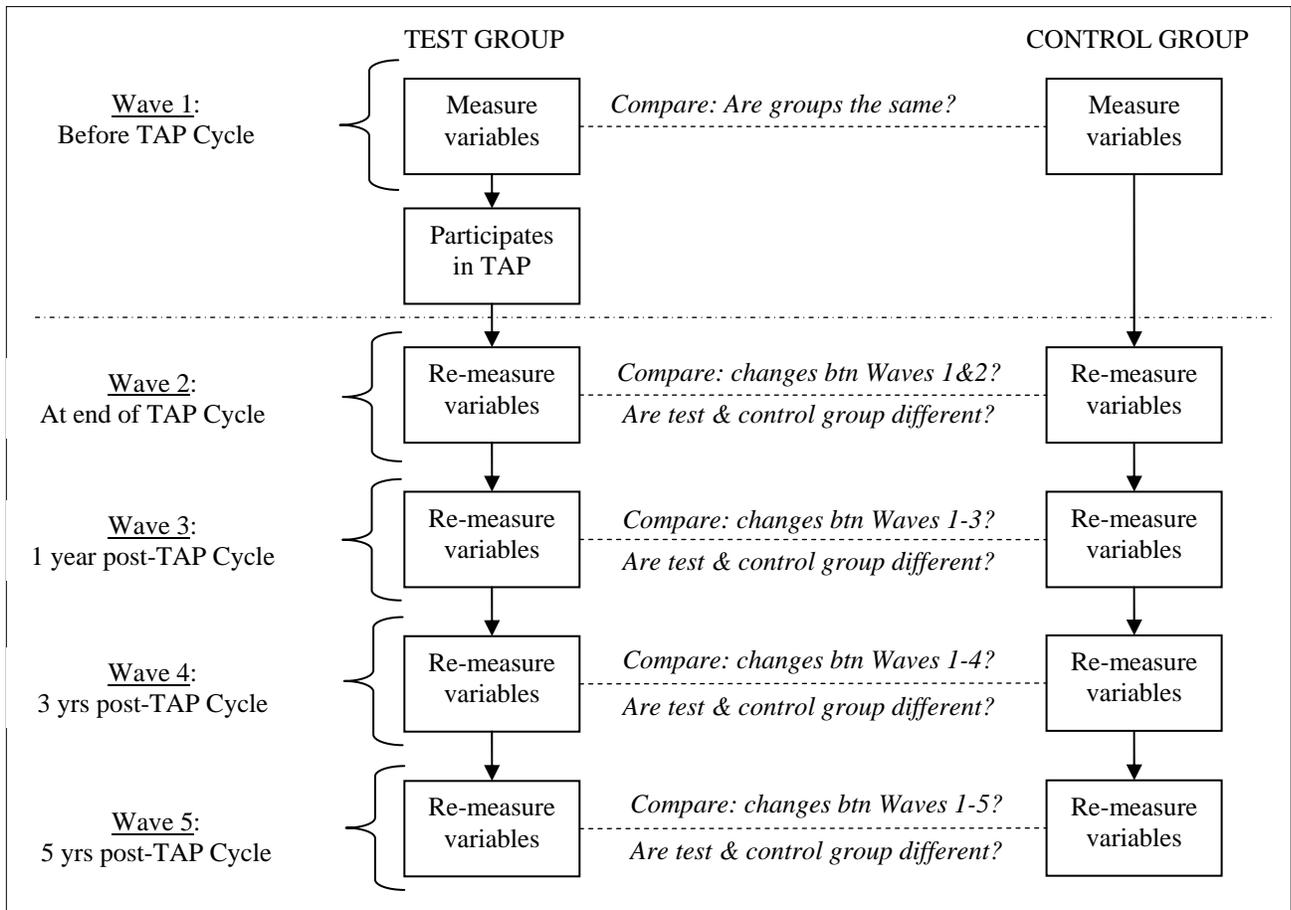


Figure 4: Longitudinal Control Group Design



Measurement of Outcomes

In order to assess the relationship between participation in TAP and the anticipated outcomes, it is necessary to define each outcome in a measurable way by choosing at least one measure for each outcome, as well as identifying data collection tools, sources and data collection methods/processes. In addition to the specific outcomes of interest, demographic data should also be gathered to explore the associations between individual characteristics (e.g. gender, race/ethnicity, age) and the measured outcomes. The short-term and medium-term outcomes of interest are laid out in TAP’s logic model. Based on these outcomes, we have created a list of recommended measures, potential data collection tools and processes in Table 4.

Table 4: Outcomes, Measures and Data Collection Short & Medium Term Outcomes

Outcome	Measures	Tools/Sources	Method
Short-term Outcomes			
Ability to work in a team	Level of self-reported ability to influence, negotiate with, and cooperate with team members	Revised Student Self-Report Teamwork Scale, see Appendix B.	Survey Questionnaire
Communication skills	Level of use of prosocial communication skills.	Communication Effectiveness Scale (CES-T) ³	Survey Questionnaire
Ability to manage emotions	Level of self-reported use of functional as compared to dysfunctional strategies for managing strong emotion	Regulation of Emotions Questionnaire (REQ); see Phillips and Power (2007) in Appendix B.	Survey Questionnaire
Self-efficacy	Level of belief that they can accomplish new or difficult tasks, or cope with adversity in various areas	General Self-Efficacy Scale (GSE), see Appendix B.	Survey Questionnaire
Self-regulatory efficacy	Level of participant's belief that they have the ability to self direct and follow through on plans and intentions	Self-Regulation Scale, see Appendix B.	Survey Questionnaire
Academic performance	Level of overall performance in school, as measured by GPA and school attendance	GPA; Attendance record	Administrative data from site or school
Delinquency	Incidences of delinquent acts, both in school and in criminal system involvement	Number of disciplinary actions at school (detention, suspension, etc); Number and severity of adjudicated convictions	Administrative data from site or school; Administrative data for court records
Technical skills in animation software	Knowledge and confidence in using 3D Studio	Skills assessment questionnaire designed by TAP	Survey Questionnaire
Medium-Term Outcomes			
Secure, supportive relationships	Number and quality of secure, supportive relationships	Multidimensional Scale of Perceived Social Support. See Appendix B.	Survey questionnaire
Criminal justice involvement	Convicted of a crime since involvement with TAP	Self-reported involvement with crime; Court records	Survey questionnaire and/or administrative data
Educational attainment	High school attended (earlier waves); Level of education completed	Survey question: 'Which high school are you attending?' 'What is the highest level of education you have completed?'	Survey questionnaire
Employment and income	Whether employed; Career/job title; Salary range	Survey question: 'Are you currently employed?' (y/n) 'What is your job title?' 'What is your salary range?' (<\$10,000; \$10,000 - \$19,999. etc.)	Survey questionnaire

³ CES-T is currently written as a 6-item scale that is completed by a young person's teacher. It could be adapted to be a self-report measure, but has not been tested for reliability and validity in this form. CES-T may be obtained by request from its authors: Jean E. Dumas (University of Geneva), Elaine A. Blechman (University of Colorado at Boulder), and Ronald Prinz (University of South Carolina).

While some outcomes, such as academic performance, are fairly simple to translate into measures, others need careful consideration of what exact outcome TAP anticipates. For example, the outcome ‘delinquency’ may be defined and operationalized in multiple ways—delinquent behavior at school, such as fighting or cutting class, which may be measured by examining the school records of disciplinary action. Delinquency may also be defined in terms of involvement with the criminal justice system as measured by arrests or adjudicated convictions⁴. Both of these measures define delinquency in relation to an institution where administrative data might be available. Alternatively, delinquency could also be defined as involvement with drug use, tagging, or other behaviors that may or may not be picked up by the school system or criminal justice system, but may be measured through self-report in a survey questionnaire. TAP may choose to measure this outcome through all of the aforementioned measures, or to concentrate more specifically on one aspect of the outcome. As the participants’ characteristics differ by site, the outcomes and their measures may need to be altered slightly at some sites. For example, measuring academic performance through the measure of school attendance may be problematic and/or less relevant for participants at the CASES site due to academic or bureaucratic processes that may hinder justice system involved students from returning to school. In this situation, alternative measures may be needed.

One of TAP’s anticipated outcomes is increased technical skills in animation software. In order to assess technical skills, particularly in the context of teaching skills, many programs use some form of an exam or test. However, TAP may not wish to use this method as using an exam format to measure technical skills may increase anxiety among participants and be overly time or resource consuming. As an alternative to this method, we recommend creating a brief questionnaire which would list the technical skills that TAP is aiming to teach and ask the participant to self-evaluate their level of proficiency (see Table 5 for examples). This type of format may be use to assess technical skills in terms of the animation process as well as skills related to learning 3D Studio in a neutral way that may serve to provide a useful skills assessment at both the beginning and the end of the cycle without the anxiety of an exam.

Table 5: Example of Self-evaluated Skill Level

For each of the following please check off the box that best answers whether you know how to...	<i>I know exactly how to do this</i>	<i>I think I know how, but might need some help</i>	<i>I'm kind of familiar, but would need a lot of help</i>	<i>No idea</i>
...create a storyboard	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
...create a bouncing ball using 3D Studio	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
...add sound to an animated object	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

⁴ We strongly recommend using convictions as a measure as arrests may be circumstantial rather than due to actual acts of delinquency (e.g., being in the wrong place at the wrong time, or living in highly policed communities).

Data Collection and Analysis

For TAP's short-term and medium-term outcome evaluations we recommend relying on pre-existing administrative databases (such as school and court records), TAP program data (e.g., intake forms, records on program participation), and using survey questionnaires for the collection of additional data. In order to gain access to individual level administrative data collected by others (e.g., schools) it is generally necessary to obtain both a release from the participant and his/her legal guardian if the participant is underage and sign a memorandum of understanding regarding the use, retention and destruction of the data with the agency collecting the data. Information regarding adjudicated convictions in New York is available through either Criminal Justice Agency and/or the Department of Corrections. Alternatively, measures to gather this information could be incorporated into a survey questionnaire. However, administrative data is frequently considered to be a more objective source of information, particularly when the questions involve potentially negative behavior.

Creating and carefully maintaining standardized program forms for each site are an excellent way to gather data. Intake forms may be used to obtain demographic data (e.g., age, race/ethnicity, socio-economic status, etc.), information on other participant characteristics (e.g. risk factors, household composition, etc.), and contact information, all of which are useful to an evaluation. In addition, forms tracking participation may be crucial to ensuring that an evaluation can take into account and control for factors such as program attendance, which may in turn influence the program's effect on participants. We recommend that TAP create an electronic database to maintain information from these forms. There are multiple software programs available for this purpose (e.g. MS Access, Filemaker). As long as the software is able to export information to MS Excel, the data may be easily imported into a statistical software package for evaluation purposes.

Survey questionnaires are a relatively cheap and convenient method of collecting a broad amount of information in a relatively non-threatening way. The questionnaires may be administered either in the traditional pencil and paper method or through a computer-based survey using an interface such as SurveyMonkey. While some potential questions and instruments have been included in the previous tables, a cohesive and streamlined survey instrument would need to be created specifically to the outcome evaluation and customized for each site. Ideally, the survey would collect information necessary to examining both the short-term and medium-term outcomes.

Individual level data obtained from TAP database, data from external administrative databases and data from survey questionnaires will need to be linked together (merged) in a program such as Excel, SPSS, or another software package capable of statistical analysis (data from paper and pencil surveys will need to be entered). We strongly recommend analysis on the

outcomes evaluation be conducted by an evaluator with a strong grounding in statistics, as the data will need to be examined using multivariate analyses in order to establish causality.

Finally, similar to the process evaluation, it is important to take measures to address ethical issues raised by some of the research methods recommended. Participation in any evaluation should always be prefaced by a consent process, by which the participant and his/her legal guardian are given the following information: (1) the purpose of the evaluation, (2) what participants will be asked to do, (3) any potential risks and benefits to the participant as well as any compensation involved, (4) what steps are being taken to ensure confidentiality of their information, (5) releases granting TAP permission to obtain information such as school records, and (6) participation is completely voluntary and they may choose to discontinue participation at any point during the study. As mentioned earlier, some of the methods recommended, such as providing information on the questionnaires, may be tedious and/or uncomfortable. Again, these risks can be minimized by making the surveys anonymous and voluntary, informing youth that they can skip any questions they feel uncomfortable answering, and providing some form of compensation (such as gift cards) for their time.

Summary and Conclusions

Vera's evaluability assessment has found that TAP has developed a coherent program design. It serves the population for whom it was designed and possesses the staff and resources called for in its design. The assessment has also determined that TAP is generally implementing its activities, though not necessarily all activities at all sites. Based on this information, process and outcome evaluations of TAP are appropriate at this time.

We recommend a three-step process evaluation which consists of a fidelity assessment, an assessment of variation in program activities, and a qualitative analysis of any discrepancies found in implementation. The first phase—fidelity assessment—needs to be completed to ensure that an outcomes evaluation will produce results that may meaningfully assess whether the program model is efficacious. Nonetheless, each successive phase will lead the outcomes evaluation to have greater explanatory power in terms of the links between program activities and outcomes and about the population that may benefit most from TAP's programming.

Concurrent to the process evaluation, we recommend that TAP conduct an outcome evaluation to measure short-term outcomes using a pretest-posttest design and medium-term outcomes using a longitudinal study design. If TAP's resources are limited, focusing on an evaluation of the short-term outcomes will provide the best return of information for money invested. We recommend that both studies incorporate a control group to provide the greatest degree of confidence that any changes found in client outcomes are the result of participation in TAP rather than contingent factors or the natural development of youth.

The key next steps that TAP should take to support an outcome evaluation are as follows. First, the program should decide whether to implement a process evaluation solely, or concurrent to an outcome evaluation. Based on this decision, TAP will need to create instruments for data collection, such as surveys, questionnaires, and observational tools. In addition, TAP should create a set of standardized program forms for gathering information at intake as well as recording information about client participation during the cycle. Finally, a program database should be created to store information from TAP's forms and any data collection that TAP decides to undertake in-house.

Evaluations may be performed either in-house (i.e., by TAP staff) or by an independent third party. In-house evaluations are more convenient and cost less, but will place additional burdens and responsibilities on program staff and are likely to be viewed as less credible than third-party evaluation. Also, certain data collection methods may not be available to an in-house evaluator. For example, observational methods will be subject to bias, and surveys or interviews with program participants about effective and ineffective methods are unlikely to produce robust data, as participants will be socially motivated to withhold criticism of the program to one of its staff members. With appropriate measures taken to reduce bias, in-house evaluations can nonetheless produce meaningful results that can inform staff members of many salient issues in program fidelity and implementation. It is also possible to take a mixed approach where an independent party is responsible for tasks that require expertise in research methods or that are subject to bias when performed by program staff (e.g., instrument creation, observation, interviews), but other techniques (such as content analyses and data collection and/or analysis of certain survey or administrative information) are the responsibility of program staff.

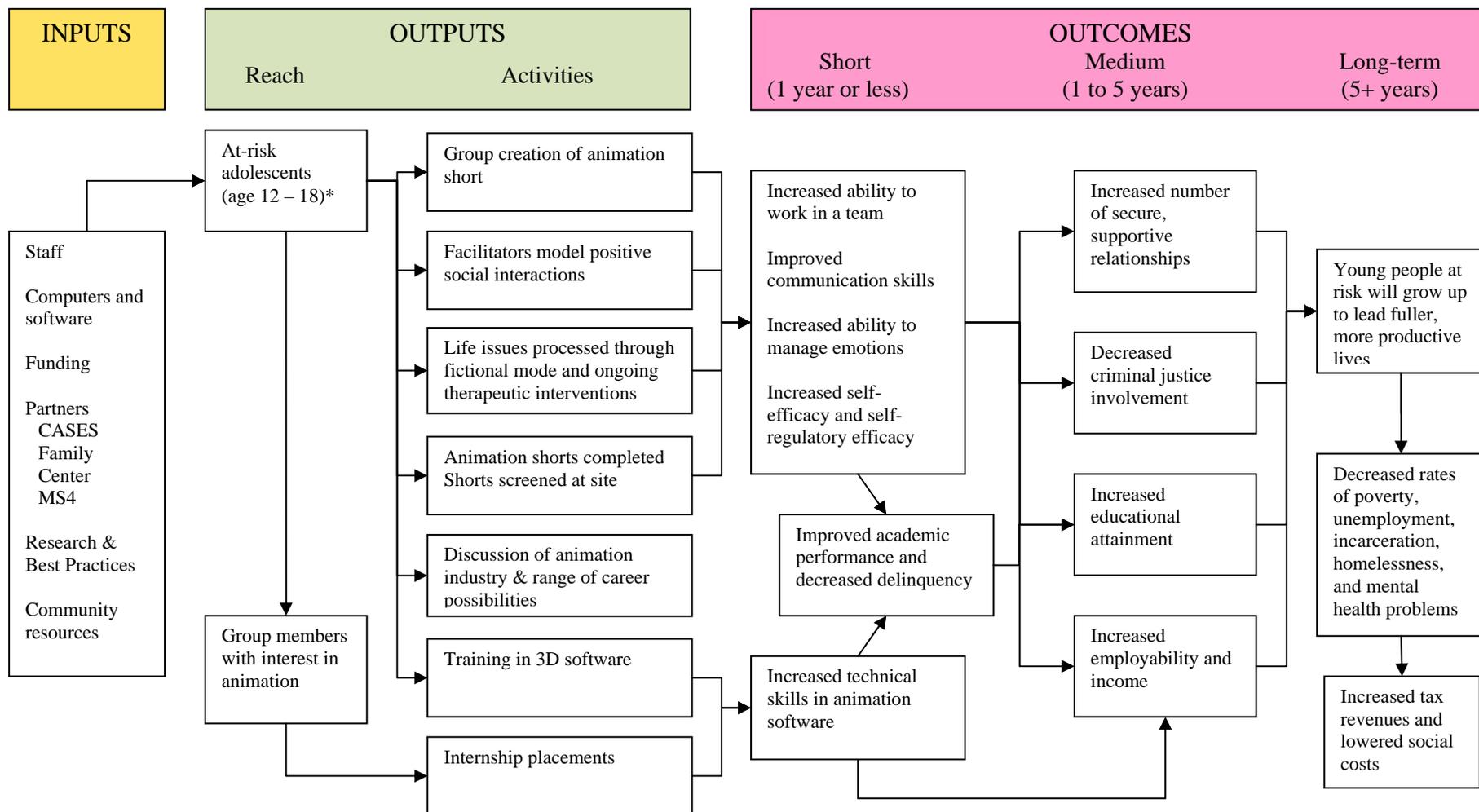
Summary of Key Next Steps for Evaluation:

1. *Decide whether to pursue: (a) process evaluation, (b) process & outcome evaluation concurrently, and which sites to evaluate.*
 - a. *If pursuing an outcome evaluation, decide whether you wish to evaluate short-term outcomes only or to include measures for medium-term outcomes as well.*
2. *Develop data collection instruments:*
 - a. *Create standardized intake and participation forms for TAP groups (consent is not necessary for these documents);*
 - b. *Develop instruments to collect information on variables of interest, including questionnaires, observations tools, etc.*
 - c. *Develop a program database to store information from TAP forms.*

Appendix A: The Animation Project Logic Model

THE ANIMATION PROJECT: LOGIC MODEL

The goal of The Animation Project is to propel the cognitive, emotional and social development of adolescents through capitalizing on an appetite for emerging technology to generate a 21st Century-ready workforce.



Appendix B: Sample Scales and Instruments

Process Evaluation:

Policy Study Associates, Inc (2005). OST Observation Instrument Domain Item Ratings. *Out of School Time (OST) Observation Instrument*. 2nd Ed. (pgs. 14, 15). Retrieved October 5, 2009 from <http://www.afterschoolresources.org/kernel/images/psaost.pdf>

Policy Study Associates, Inc. (n.d.) *Evaluation of the TASC After-School Program High School Student Survey – School Year 2001/2002*. (p. 10). Retrieved October 5, 2009 from <http://www.policystudies.com/studies/youth/Year%204%20High%20School%20Student%20Survey.pdf>

Outcome Evaluation:

Phillips, K.F.V., & Power, M.J. (2007). A new self report measure of emotion regulation in adolescents: The regulation of emotions questionnaire. *Clinical Psychology and Psychotherapy*, 14, 145-156.

Zhuang, X., MacCann, C., Wang, L., Liu, L., & Roberts, R.D. (2008). *Development and Validity Evidence Supporting a Teamwork and Collaboration Assessment for High School Students*. (p.18). ETS RR-08-50. Educational Testing Service: Princeton, NJ.

Zimet, G.D., Dahlem, N.W., Zimet, S.G. & Farley, G.K. (1988). The Multidimensional Scale of Perceived Social Support. *Journal of Personality Assessment*, 52, 30-41.

Freien Universität Berlin (n.d.). *The General Self-Efficacy Scale (GSE)*. Retrieved October 5, 2009, from <http://userpage.fu-berlin.de/~health/engscal.htm>

Schwarzer, R., Diehl, M., & Schmitz, G.S. (1999). *Self-Regulation Scale*. Retrieved October 5, 2009, from http://userpage.fu-berlin.de/~health/selfreg_e.htm

Appendix B: Sample Scales and Instruments

Policy Study Associates, Inc (2005). OST Observation Instrument Domain Item Ratings. *Out of School Time (OST) Observation Instrument*. 2nd Ed. (pgs. 14, 15). Retrieved October 5, 2009 from <http://www.afterschoolresources.org/kernel/images/psaost.pdf>

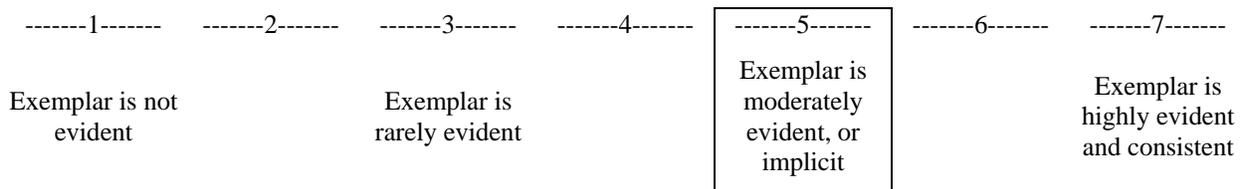
OST OBSERVATION INSTRUMENT DOMAIN ITEM RATINGS

The following are five youth development domains that include four to seven indicators of youth development practices within each. The purpose of the observation instrument is to measure the extent to which these indicators – also called “items” – may or may not be present in each 15-minute observation segment. necessarily negative.

After 15 minutes of observation, assign a rating of 1 (not evident) to 7 (highly evident and consistent) for each item below. To select a rating, first move to the ODD NUMBER that most closely reflects how evident and pervasive an item is. If that number does not precisely reflect the level of evidence observed, then move down or up to the adjacent even number that more accurately reflects the item’s level of presence within an activity. Note that each item/indicator may not be present/applicable in each observation; therefore, a rating of “1” is not necessarily negative.

The “5” rating is also used in cases where the exemplar’s presence is implicit within the activity. For instance, if youth are generally friendly to each other throughout the observation, but most do not go beyond the casual, friendly interaction, the rating would be a “5”.

RATINGS:



RELATIONSHIP BUILDING : <u>all or most YOUTH</u>		PARTICIPATION: <u>all or most YOUTH</u>	
A	Are friendly and relaxed with one another. Youth socialize informally. They are relaxed in their interactions with each other. They appear to enjoy one another’s company.	F	Are on-task. Youth are focused, attentive, and not easily distracted from the task/project. They follow along with the staff and/or follow directions to carry-on an individual or group task.
B	Respect one another. Youth refrain from causing disruptions that interfere with others accomplishing their own tasks. When working together, they consider one another’s viewpoints. They refrain from derogatory comments or actions about the individual person and the work s/he is doing; if disagreements occur, they are handled constructively.	G	Listen actively and attentively to peers and staff. Youth listen and respond to each other and staff. They appear interested in what others have to say. They look at peers and/or staff when they speak, and they provide concrete and constructive feedback about ideas or actions.
C	Show positive affect to staff. Youth interact with the staff, and these interactions are generally friendly interactions. For example, they may smile at staff, laugh with them, and/or share good-natured jokes.	H	Contribute opinions, ideas and/or concerns to discussions. Youth discuss/express their ideas and respond to staff questions and/or spontaneously share connections they’ve made. This item goes beyond basic Q&A and refers to sharing that is part of the activity and within the class norms. Calling out – or disruptively talking out of turn – is not part of this item.
D	Assist one another. One or more youth formally or informally reach out to help/mentor peers and help them think about and figure out how to complete a task. This item refers to assistance that is intentional and prolonged, going beyond answering an incidental question.	I	Have opportunities to make <u>meaningful choices</u>. <i>Within this activity</i> , youth choose what they do, how they do it, and/or with whom they collaborate, and they experience the consequences of their choices. This item refers to genuine options, not simple choices such as choosing between two types of games, or two sets of homework pages.
E	Are collaborative. Youth work together/share materials to accomplish tasks. This item is different from item D (above), as in collaboration, youth are equal partners in the work (rather than one student assisting/mentoring/tutoring another). This item can include working together on assigned teams, if youth are working together to get a better result.	J	Take leadership responsibility/roles. Youth have meaningful responsibility for directing, mentoring or assisting one another to achieve an outcome; they lead some part of the activity by organizing a task or a whole activity, or by leading a group of youth within the activity.

RELATIONSHIP BUILDING: with all youth, <u>STAFF...</u>		INSTRUCTIONAL STRATEGIES: <u>STAFF...</u>	
K	Use positive behavior management techniques that allow for youth to accomplish the activity's objectives. They set consistent limits and communicate clear expectations for behavioral standards, and these are appropriate to the age of the youth and the activity type. When disciplining youth, they do so in a firm manner, without unnecessary accusations, threats, or anger.	R	Communicate goals, purpose, expectations. Staff make clear the value and purpose of what youth are doing <u>and/or</u> what they expect them to accomplish. This item goes beyond how youth are expected to behave (which would be captured in item K).
L	Are equitable and inclusive. Youth are provided equal opportunity to participate in an activity and are rewarded/disciplined similarly for like actions. Staff encourage the participation of all youth, regardless of gender, race, language ability, or other evident differences among students. They try to engage students who appear isolated; they do not appear to favor a particular student or small cluster of students.	S	Verbally recognize youth's efforts and accomplishments. Staff acknowledge participation and progress in order to encourage youth.
M	Show positive affect toward youth. Staff interact with youth, and these interactions are generally friendly. For example, their tone is caring, and/or they use positive language, smile, laugh, or share good-natured jokes.	T	Assist youth without taking control. Staff may coach, demonstrate, or employ scaffolding techniques that help youth to gain a better understanding of a concept or complete an action on their own. Staff refrain from taking over a task or doing something on behalf of the youth. This assistance goes beyond checking that work is completed.
N	Attentively listen to and/or observe youth. Staff look at youth when they speak and acknowledge what youth have said by responding and/or reacting. They pay attention to youth as they complete a task and appear interested in what they are saying/doing.	U	Ask youth to expand upon their answers and ideas. Staff encourage youth to explain their answers, evidence, or conclusions. They may ask youth 'why', 'how' and 'if' questions to get them to expand, explore, better clarify, articulate, or concretize their thoughts/ideas. This item goes beyond staff-elicited Q&A.
O	Encourage youth to share their ideas, opinions and concerns. Staff <u>actively elicit</u> youth ideas, opinions and concerns through discussion and/or writing. This item goes beyond basic Q&A.	V	Challenge youth to move beyond their current level of competency. Staff give constructive feedback that is meant to help youth to gauge their progress. Staff help youth determine ways to push themselves intellectually, creatively, and/or physically.
P	Engage personally with youth. Staff show interest in youth as individuals, ask about youth's interests, and engage about events in their lives.	W	Employ varied teaching strategies. In order to engage students and/or reach those with different learning styles, staff diversify instructional strategies, which may include the use of two or more of the following: direct instruction, coaching, modeling, demonstrating, or others. Varied instructional strategies can occur simultaneously <u>and/or</u> sequentially within the observation period. This item does not include coupling a staff-directed instruction with youth working together, as described above.
Q	Guide for positive peer interactions. Staff intentionally encourage positive interactions and/or directly teach interpersonal skills. They teach these skills through planned activity content or through intervening constructively and calmly to address bullying or teasing behavior, redirecting youth and/or explaining or discussing why negative behavior is unacceptable. This item does not refer to behavior management, as described above (see item K).	X	Plan for/ask youth to work together. Staff plan for and/or ask youth to work together, solve problems, and/or accomplish tasks. The focus of the activity is youth to youth, rather than youth to staff. This item goes beyond staff assigned teams for competitive games and sports. In the case of staff assigned teams, staff would also need to be directing youth to collaborate, plan, devise, etc., in order for this item to be rated as staff asking youth to work together.

Appendix B: Sample Scales and Instruments

Policy Study Associates, Inc. (n.d.) *Evaluation of the TASC After-School Program High School Student Survey – School Year 2001/2002*. (p. 10). Retrieved October 5, 2009 from <http://www.policystudies.com/studies/youth/Year%204%20High%20School%20Student%20Survey.pdf>

19. The following are statements that might describe how you feel about the teachers at the AFTER-SCHOOL PROGRAM. For each statement, please circle whether you agree a lot, agree a little, disagree a little, disagree a lot, or have no opinion.

	<i>Agree a lot</i>	<i>Agree a little</i>	<i>Disagree a little</i>	<i>Disagree a lot</i>	<i>No opinion</i>
a. Teachers and students treat each other with respect in this program	1	2	3	4	5
b. I feel that I can talk to the teachers in this program about things that are bothering me	1	2	3	4	5
c. Teachers and students in this program don't seem to like each other	1	2	3	4	5
d. The teachers in this program really care about me	1	2	3	4	5
e. The teachers in this program always keep their promises	1	2	3	4	5
f. The teachers in this program don't care what I think	1	2	3	4	5
g. The teachers in this program always try to be fair	1	2	3	4	5
h. Teachers in this program punish kids without even knowing what really happened	1	2	3	4	5
i. I feel safe and comfortable with the teachers in this program	1	2	3	4	5

Appendix B: Sample Scales and Instruments

Phillips, K.F.V., & Power, M.J. (2007). A new self report measure of emotion regulation in adolescents: The regulation of emotions questionnaire. *Clinical Psychology and Psychotherapy*, *14*, 145-156.

Assessment *A New Self-Report Measure of Emotion Regulation in Adolescents: The Regulation of Emotions Questionnaire*

K. F. V. Phillips and M. J. Power*

Section of Clinical & Health Psychology, University of Edinburgh, Edinburgh, UK

Emotion regulation is an emerging field in psychology concerned with the processes involved in recognizing, monitoring, evaluating and modifying emotional reactions. Few measures have been developed specifically to assess individual differences in the use of emotion regulation strategies. This paper reports on a new self-report instrument designed to assess the frequency with which adolescents use both functional and dysfunctional emotion regulation strategies, which draw on both internal (intrapersonal) and external (interpersonal) resources to regulate emotions. The new measure was used in a cross-sectional survey of 225 adolescents. Factor analyses supported the structure of the measure, which has four scales: 'internal-functional', 'internal-dysfunctional', 'external-functional' and 'external-dysfunctional' emotion regulation. The new measure demonstrated good psychometric properties. Frequent use of dysfunctional emotion regulation was significantly positively correlated with parent reports of level of emotional and behavioural problems, and also with self-reported level of psychosomatic health problems. Frequent use of functional emotion regulation was significantly positively correlated with self-reported quality of life across a number of different dimensions. These results provide evidence for the validity of the measure. The results also highlight that assessment of both functional and dysfunctional emotion regulation is important. Copyright © 2007 John Wiley & Sons, Ltd.

INTRODUCTION

Emotion Regulation

Contemporary theories of emotion generally suggest that emotions exist as part of the rational human element, rather than as 'passion at odds

*Correspondence to: Professor Mick Power, Section of Clinical and Health Psychology, University of Edinburgh, Medical School, Teviot Place, Edinburgh EH8 9AG, UK.
E-mail: mjpower@staffmail.ed.ac.uk

with reason' as Plato argued. As a result, it is now accepted that there is scope for individuals to regulate their own emotional experience. Research into emotion regulation has emerged from the psychoanalytic tradition and the 'stress and coping' field (Gross, 1999). With regard to the latter, Lazarus and Folkman's (1984) concept of 'emotion-focused coping' has laid the groundwork for much of the contemporary study of emotion regulation (e.g., Gross, 1999; Losoya, Eisenberg, & Fabes, 1998).

Gross (1999) proposes that 'emotion regulation' can be categorized as a subcategory of the broader category of 'coping'. Coping is considered the broader category (Gross, 1999) because unlike emotion regulation, it can include taking actions that are non-emotional to achieve goals that are also non-emotional (e.g., Scheier, Weintraub, & Carver, 1986). In contrast, emotion regulation refers specifically to the actions taken to achieve emotional goals. Thompson (1994) suggests that the term 'emotion regulation' refers to the processes, both intrinsic and extrinsic, that are responsible for learning to recognize, monitor, evaluate and modify emotional reactions. Both positive and negative emotions are regulated (Gross, 1998). The processes that are involved may be automatic or controlled, conscious or unconscious, and can come into effect at different points in the emotion generation process (Gross, 1998).

Gross (1998) argues that it is better not to judge *a priori* whether particular emotion regulation strategies are 'good' or 'bad', suggesting that this will enable the field to avoid the problems encountered in the stress and coping literature where coping is viewed as adaptive, and defences as maladaptive. It is probable that different emotion regulation strategies will be functional or adaptive in some contexts and not others (e.g., Southam-Gerow & Kendall, 2002). However, it may also be true that individuals have tendencies towards using particular types of regulatory strategies, resulting in an emotion regulatory style that is generally functional or dysfunctional. Indeed, John and Gross (2004) found evidence for 'healthy' and 'unhealthy' types of emotion regulation. Therefore, to measure emotion regulation comprehensively, it is important to assess both adaptive and maladaptive emotion regulation.

Emotion regulation is linked to aspects of functioning. Kostiuk and Fouts (2002) suggest that individuals who lack the ability to regulate emotions successfully misidentify and misdirect their emotional experience, thus hindering their ability

to function adaptively and appropriately. Gross (1999) points to links between emotion regulation and psychological problems, suggesting that dysfunctional emotion regulation is implicated in over half of the DSM-IV Axis I disorders and in all of the Axis II disorders (American Psychiatric Association, 1994). Emotion regulation competencies may also influence physical health. Indeed, emotional factors have been implicated in physical illness ever since it has been understood that emotion generation is linked to arousal of the autonomic nervous system. For instance, the inhibition and excessive expression of emotion has been linked to illness (e.g., Pennebaker, 1990; Siegeman, 1993). Gross and Munoz (1995) suggest that individual differences in emotion regulation can aid or disrupt the capacity for work, social interaction and overall satisfaction with life. Measures of emotion regulation should demonstrate expected relationships with these factors.

Emotion Regulation in Young People

As discussed previously, the study of emotion regulation stems from work on emotion-focused coping. Studies with young people tend to show that the use of emotion-focused coping is linked to poorer outcomes than problem-focused coping. This may be taken as evidence to suggest that attempts to focus on and cope with (or regulate) emotions are maladaptive, especially when compared to managing the problems responsible for eliciting emotions. However, these results reflect that disengagement with stressors or emotions is problematic, rather than that a focus on and regulation of emotions is unhelpful *per se*.

The Assessment of Individual Differences in Emotion Regulation

Southam-Gerow and Kendall (2002) argue that further research into emotion regulation may facilitate a better understanding of psychopathology in young people. However, measures of emotion regulation developed specifically for use with adolescents are scarce. Numerous measures of coping have been developed for and used with adolescents. However, since coping differs from emotion regulation (e.g., Gross, 1999), these measures often do not focus on how young people respond to their emotions. Indeed, these measures tend to ask young people to consider strategies used during times of stress and therefore may miss

strategies used in everyday situations to regulate all their emotions, as opposed to those elicited according to a particular stressor.

Some measures aim to specifically tap into aspects of emotion regulation. For example, Garnefski, Kraaij, and Spinhoven (2001) developed the Cognitive Emotion Regulation Questionnaire. This measure the theoretically adaptive or functional strategies and the theoretically non-adaptive or dysfunctional strategies. However, the measure focuses only on cognitive strategies used to regulate emotions. Therefore, other types of strategies, such as behavioural strategies, are missed. Moreover, the measure has been developed for and has been used mostly with adults (e.g., Garnefski et al., 2001), though it has been used to assess the relationship between cognitive emotion regulation strategies, life events and depressive symptomatology in adolescents (Garnefski et al., 2003).

The Emotion Regulation Questionnaire (ERQ, Gross & John, 2003) is another measure that focuses specifically on emotion regulation. Gross and John (2003) acknowledge that the measure focuses on a small number of strategies. Indeed, the ERQ assesses only positive reappraisal and suppression. Positive reappraisal is considered an adaptive emotion regulation strategy, and suppression a maladaptive strategy. This measure is reliable and valid (Gross & John, 2003). However, it has been developed and has been used in studies with adults, and therefore it is not certain whether the measure would be suitable for use with young people.

Another measure of emotion regulation (or dysregulation) has been developed by Gratz and Roemer (2004) and is called the Difficulties in Emotion Regulation Scale. This measure assesses many facets of emotion regulation, including 'lack of emotional awareness', 'non-acceptance of emotional responses' and 'limited access to emotion regulation strategies' (Gratz & Roemer, 2004). It has a strong focus on dysfunctional emotion regulation, though some items refer to adaptive strategies but are then reverse scored. This measure has also been developed and used with adults; thus, suitability for use with young people has not been confirmed.

Measures that specifically aim to measure emotion regulation in young people do exist. However, a number of these are designed to be completed by parents and, therefore, may be more suited to measuring emotion regulation in young children rather than adolescents. For example, the Emotion Regulation Checklist (ERC, Shields & Cicchetti, 1997) is an adult-report measure of a child's emotion regulation. It measures the extent to

which children are able to regulate their emotions with items such as 'can modulate excitement (doesn't get carried away in high energy situations)'. Although the ERC shows good psychometric properties and has been shown to discriminate between well-regulated and dysregulated children, it appears to tap into the outcomes of using dysfunctional emotion regulation strategies as opposed to the strategies that a child uses or does not use.

It is considered preferable to have multiple respondents when assessing emotion regulation variables. Therefore, there is certainly a need for proxy measures. Additionally, these measures are likely to be particularly useful when assessing the emotion regulation of very young children, since the children may be less able to report on their own use of emotion regulation strategies. However, measures that specifically ask for an adolescent's view of their style of emotion regulation are also needed. Parents may be subject to certain biases, such as the desire to present their child in a positive light. Moreover, parents are likely to be unaware of the personal strategies that young people may use to regulate their emotions. Therefore, a self-report measure of adolescent emotion regulation is needed. Indeed, adolescence is a time of increasing cognitive complexity; thus, adolescents are likely to be able to reflect on certain aspects of their own emotion regulation.

A New Measure of Emotion Regulation in Adolescents

A self-report measure of emotion regulation for adolescents is needed. Such a measure should be based on relevant theories concerning emotion regulation processes, as well as what constitutes functional and dysfunctional emotion regulation. Gross (1998) describes a process model of emotion regulation that outlines points at which emotions can be regulated: antecedent strategies (those that occur prior to the generation of emotion) and response-focused strategies (those that are used to manage emotions after they have been generated). Gross and John (2003) used this framework to assess two types of emotion regulation: cognitive reappraisal and expressive suppression. Cognitive reappraisal is suggested to be an antecedent-focused, 'healthy' strategy, whereas expressive suppression is a response-focused, 'unhealthy' strategy. Although the measure based on this model has generated interesting results in adult samples (e.g., John & Gross, 2004), it can be argued that there are helpful and unhelpful versions of

both reappraisal and expressive suppression. For example, suppression of anxiety elicited during an interview may be 'healthy', whereas an 'unhealthy' type of reappraisal may be to reappraise a situation negatively.

An alternative framework on which to base the assessment of emotion regulation is the functional account of emotions. This account proposes that the important functions of emotions are to provide useful information about situations and to enhance the individual's capacity to deal with them. For example, the experience of anger gives the individual the capacity to retaliate in some way; the experience of anxiety alerts an individual to a threat to their goals and plans, and the experience of happiness alerts individuals that they are acting consistently with their goals and plans (Power & Dalgleish, 1997). Therefore, a regulatory strategy that uses the information provided by an emotion would be adaptive or functional since it involves 'holding' and processing the emotion, thereby facilitating goal-directed behaviour that promotes well-being (e.g., Power & Dalgleish, 1997). In contrast, strategies that do not process and use the information in a helpful way, perhaps rejecting or blocking emotions instead, would be considered dysfunctional. These strategies prevent the development of tolerance to emotions (Chapman, Specht, & Celluci, 2005) and may actually lead to amplification of emotional distress due to the generation of secondary emotions, such as feeling anxious about the initial 'bad' feelings (e.g., Linehan, 1993) or through the 'coupling' of emotions (e.g., Power & Dalgleish, 1997).

Another important factor to consider when assessing emotion regulation is the nature of the resources used to regulate emotions (e.g., Eisenberg, Champion, & Ma, 2004). Consistent with existing psychological models such as attribution theory (e.g., Weiner, 1986), emotion regulation strategies can be classified as those that utilize primarily personal or 'internal' resources and those that utilize environmental or 'external' resources (i.e., using actual attachment-related resources). Using both internal and external resources can be functional (and dysfunctional), though the use of internalized resources increases with development; thus, assessing both internal and external strategies would be important in young people.

In summary, this paper reports on the development and use of a new measure of individual differences in adolescent emotion regulation. It was hypothesized that emotion regulation strategies could be classified into internal-dysfunctional,

internal-functional, external-dysfunctional and external-functional. It was also hypothesized that frequent use of dysfunctional emotion regulation strategies would be significantly positively correlated with emotional, behavioural and psychosomatic health problems in adolescents (e.g., Gross, 1999; Kostiuk & Fouts, 2002; Pennebaker, 1990). In contrast, it was predicted that dysfunctional emotion regulation would be significantly negatively correlated with quality of life (QOL), whereas functional emotion regulation would be positively correlated with QOL in adolescents (e.g., Gross & Munoz, 1995; Kostiuk & Fouts, 2002).

METHOD

Initial Development of the Regulation of Emotions Questionnaire (REQ)

The REQ was developed according to the framework described (functional vs. dysfunctional, internal vs. external), based on the literature on emotions and emotion regulation. Forty-two items were constructed to represent emotion regulation strategies. To ensure that the items corresponded with the conceptual model of emotion regulation, 10 experts (two clinical psychologists, one health psychologist, four research psychologists, one social psychologist, one occupational psychologist and one consultant child and adolescent psychiatrist) rated each of 42 suggested items. The experts were required to state whether they considered an item to represent an emotion regulation strategy at all and additionally to state whether they thought each item referred to an 'internal-functional', 'internal-dysfunctional', 'external-functional' or 'external-dysfunctional' emotion regulation strategy, or more than one of these types. The experts were also given the opportunity to reword items, suggest alternative items and offer suggestions of emotion regulation strategies that were not covered by the measure. The use of expert consultancy can be considered to add to the validity of the measure.

From the 42 suggested items, 32 items were retained on the grounds that there was a consensus regarding classification. The 32-item pilot measure comprised 8 items assessing internal-functional emotion regulation strategies, 11 internal-dysfunctional items, 6 external-functional items and 7 external-dysfunctional items. A 5-point frequency scale from 1 ('not at all') to 5 ('always') was used with each item. Examples of the items for each of the sub-scales are 'I keep the feeling locked up inside' (internal-

dysfunctional), 'I put the situation into perspective' (internal-functional), 'I try to make others feel bad' (external-dysfunctional) and 'I talk to someone about how I feel' (external-functional).

Using the Emotion Regulation Measure

Design

A cross-sectional survey design was used in this study. Adolescents were asked to complete a series of self-report measures, including the new REQ.

Participants

Two hundred and twenty-five participants aged between 12 and 19 took part in the main study (mean age = 15.06). Of these, 119 were female (53.1%), 105 were male (46.9%) and one participant did not state his or her gender. The participants came from a wide range of socioeconomic groups across the UK.

Procedure

A mail survey of adolescents living in the UK was conducted. Interviewers from a market research company recruited families by telephone, following computer generation of random UK telephone numbers. Families who agreed to take part were mailed a survey pack with consent form and a parent consent form for adolescents under the age of 16. Completed questionnaires were returned to the research team by mail in the stamped-addressed envelopes provided.

Measures

In addition to the pilot REQ, the following measures were also completed in order that expected relationships between emotion regulation and functioning could be tested.

1. Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997). This 25-item measure of emotional and behavioural problems in young people consists of five scales that assess (1) emotional symptoms, (2) conduct problems, (3) hyperactivity, (4) peer problems and (5) prosocial behaviour. The scales of the measure have demonstrated test-retest reliability of 0.70–0.85 and internal consistency of 0.51–0.76 (Goodman, 1997). The parent version was used in this study to reduce the questionnaire burden for the young people.
2. Psychosomatic Health Problems Scale (Health Behaviour in School-Age Children [HBSC]; Currie, Samdal, Boyce, & Smith, 2001). This is

an eight-item self-report scale used by the HBSC study (Currie et al., 2001). It asks young people how often they experience the following: headaches, stomachaches, backache, feeling low, irritability, anxiety, difficulties getting to sleep and dizziness. Answers are given on a five-point scale. In the current study, this scale had a Cronbach's alpha of 0.81.

3. KIDSCREEN (KIDSCREEN-52; Ravens-Sieberer et al., 2001, 2005). The KIDSCREEN-52 instrument is a 52-item self-report measure of QOL across 10 dimensions. The KIDSCREEN measure is a reliable and valid self-report measure of QOL in adolescents (e.g., Ravens-Sieberer et al., 2005), with Cronbach's alphas ranging from 0.76 to 0.89. Inclusion of this measure permitted analysis of whether the REQ scales that assess functional (adaptive) emotion regulation strategies were associated with satisfaction with well-being in different domains.

RESULTS

Refinement of the Emotion Regulation Measure

Prior to further analysis, the emotion regulation measure was refined. Items were retained or dropped based on descriptive statistics and reliability analyses. In these analyses, cases were required to have less than 20% missing values on a scale in order to be included in the analysis for that scale. Frequency analyses were carried out to determine the distribution of responses across the five-point scales for each item. Items were identified as problematic if two or more adjacent scale points showed fewer than 10% of the responses on aggregate. This analysis identified items that were likely to have skewness and kurtosis. Items were also discarded if they loaded as highly or higher on scales other than their own predicted scales. Twelve items were eliminated according to these criteria, leaving 20 items.

Factor Analysis

A principal components extraction with varimax rotation was conducted with the retained items. Two items were dropped as they had loadings of below 0.45, which Comrey and Lee (1992) suggested was 'fair'. The resulting factor analysis of 18 items had a Kaiser measure of sampling adequacy

Table 1. Item loadings on each factor of the Regulation of Emotions Questionnaire after varimax rotation

Item	Factor			
	1	2	3	4
7: I review (rethink) my thoughts or beliefs.	0.632	5.78E	-0.285	0.272
15: I review (rethink) my goals or plans	0.759	4.35E	-8.37E	0.229
18: I put the situation into perspective	0.728	-3.32E	0.136	0.209
19: I concentrate on a pleasant activity	0.582	-4.87E	0.351	6.60E
25: I plan what I could do better next time	0.737	-8.51E	3.13E	4.19E
4: I take my feelings out on others verbally (e.g., shouting, arguing)	2.34E	0.672	7.25E	-3.13E
16: I take my feelings out on others physically (e.g., fighting, lashing out)	-5.64E	0.725	0.138	8.95E
23: I try to make others feel bad (e.g., being rude, ignoring them)	1.13E	0.692	0.118	-8.88E
30: I bully other people	2.19E	0.758	0.116	-8.52E
31: I take my feelings out on objects around me (e.g., deliberately . . .)	-7.16E	0.692	0.169	0.126
8: I harm or punish myself in some way	0.241	0.172	0.580	-0.107
11: I dwell on my thoughts and feelings (e.g., it goes round and round . . .)	-2.35E	6.45E	0.768	3.70E
24: I think about people better off and make myself feel worse	-9.94E	0.203	0.571	6.42E
28: I keep the feeling locked up inside	9.31E	7.90E	0.652	0.326
32: Things feel unreal (e.g., I feel strange, things around me feel strange . . .)	-1.59E	0.193	0.675	-9.05E
2: I talk to someone about how I feel	0.200	-2.37E	0.201	0.829
6: I ask others for advice	0.153	-2.01E	-0.158	0.732
14: I seek physical contact from friends or family (e.g., a hug, hold hands)	0.371	0.135	0.136	0.760

of 0.772, which is still well above the expected cut-off of 0.6, indicating good factorability (Tabachnick & Fidell, 2001). Four factors had eigenvalues above 1, and these factors explained 56% of the variance.

Table 1 shows that the items loading on factor 1 represent the internal-functional scale; the items loading on factor 2 represent the external-dysfunctional scale; the items loading on factor 3 represent the internal-dysfunctional scale, and those loading on factor 4 represent the external-dysfunctional scale. All factors had loadings of above 0.55 (30% overlapping variance) on their own scales. Comrey and Lee (1992) suggested that loadings in excess of 0.55 were 'good'. Moreover, most of the item loadings were actually in excess of 0.63 or 0.71, which Comrey and Lee (1992) suggested were 'very good' and 'excellent', respectively.

Structural Equation Modelling

Structural equation modelling with the EQS programme (Bentler, 1989) was used to test the original hypothesized four-factor model. It was not possible to perform this analysis using the split-half method due to insufficient numbers, though it is acknowledged that this would certainly be preferable. The five items that performed best in the previous analyses for each of the scales were selected for analysis. However, an item from the external-functional scale that did not perform well

in the exploratory factor analysis was included since a three-item scale precludes assessment of a scale's internal consistency (e.g., Tabachnick & Fidell, 2001).

A correlated factor model, which allowed correlations between conceptually related variables, proved to be a highly significant improvement on the original model that constrained the factor correlations to be 0. The independence chi-square ($\chi^2 = 1259.34$, degree of freedom [df] = 171, $p < 0.001$) indicated that there was a relationship between the variables. The standardized residuals were fairly low, with the average absolute standardized residual equalling 0.05. All path coefficients between measured variables and factors in the model were significant ($p < 0.05$). The results show that the comparative fit index was 0.905, which is greater than the suggested value of 0.9, which indicates an acceptable model fit (Tabachnick & Fidell, 2001). Figure 1 below shows the path diagram of the model.

Figure 1 shows that the overall fit of the model improves if F1 (internal-dysfunctional) is allowed to correlate with F3 (external-dysfunctional), and F2 (internal-functional) is allowed to correlate with F4 (external-functional).

Descriptive Statistics

Table 2 shows the means, standard deviation and Cronbach's alpha for each scale. Table 2 shows that

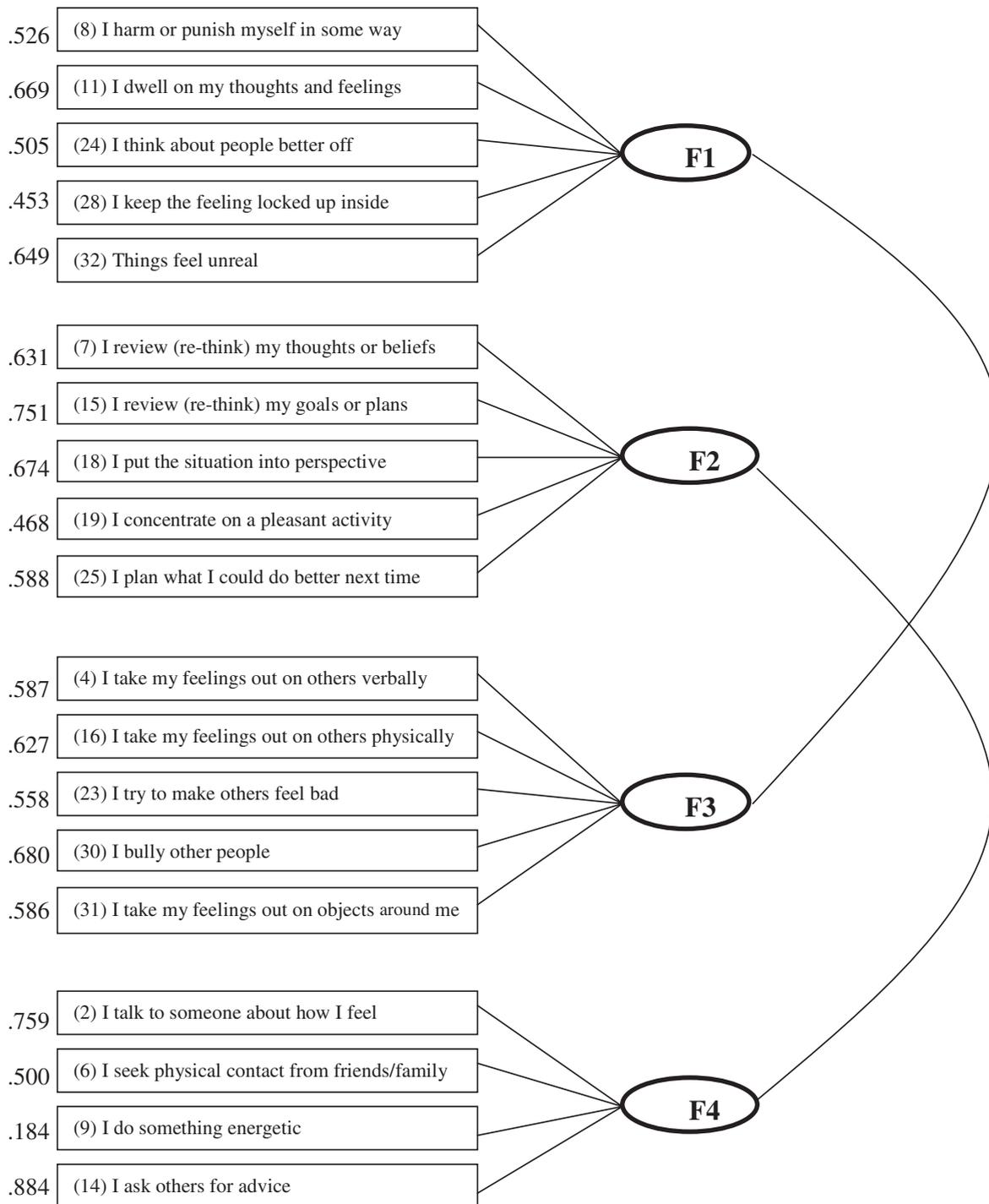


Figure 1. Path diagram of the 19-item Regulation of Emotions Questionnaire

Table 2. Mean, standard deviation (SD) and internal consistency of each Regulation of Emotions Questionnaire scale

Scale	Mean	SD	Cronbach's alpha
Internal-dysfunctional ER	2.07	0.70	0.72
Internal-functional ER	2.89	0.70	0.76
External-dysfunctional ER	1.55	0.53	0.76
External-functional ER	2.83	0.79	0.66

ER = emotion regulation.

Table 3. Correlations between scores on the four emotion regulation (ER) scales (Regulation of Emotions Questionnaire) and scores on the Strengths and Difficulties Questionnaire (SDQ) scales

	Internal-dysfunctional ER	Internal-functional ER	External-dysfunctional ER	External-functional ER
SDQ: emotional symptoms	0.38**	-0.11	0.18**	-0.04
SDQ: conduct problems	0.21**	-0.05	0.43**	-0.04
SDQ: hyperactivity	0.18**	-0.11	0.29**	-0.05
SDQ: peer problems	0.24**	-0.12	0.24**	-0.20**
SDQ: pro-social behaviour	-0.18**	0.07	-0.27**	0.18**
SDQ: total difficulties	0.35**	-0.14*	0.39**	-0.11

*Correlation is significant at the 0.05 level. **Correlation is significant at the 0.01 level.

all the scales demonstrated acceptable internal consistency.

Validity of the Measure: Relationships between Emotion Regulation and (1) Emotional and Behavioural Problems, (2) Psychosomatic Health Problems and (3) QOL

Pearson's correlations were used to examine the associations between scores on the four emotion regulation strategy scales from the REQ and scores on the SDQ, psychosomatic health problems scale and KIDSCREEN.

REQ and SDQ

Table 3 shows that both the internal-dysfunctional and external-dysfunctional emotion regulation scales correlated significantly positively with every scale of the SDQ, except for the pro-social scale. This shows that the more frequently internal- and external-dysfunctional strategies are used, the greater the level of emotional and behavioural problems. In contrast, frequent use of internal- and external-dysfunctional emotion regulation strategies was associated with less pro-social behaviour. The functional emotion regulation strategies showed few associations with the SDQ scales, though external-functional regulation was significantly negatively correlated with the peer problems scale and significantly positively correlated

with the pro-social behaviour scale. These results support the hypotheses based on past literature and therefore provide support for the validity of the new instrument.

REQ and Psychosomatic Health Problems

Table 4 shows that there was a highly significant positive correlation between the reported use of internal-dysfunctional emotion regulation strategies and the reported experience of psychosomatic health problems. In addition, there was a significant positive correlation between the use of external-dysfunctional emotion regulation strategies and psychosomatic health problems. As expected, the use of internal and external-functional regulatory strategies were not related to psychosomatic health problems.

REQ and KIDSCREEN

Table 5 shows that there were significant associations between the strategies adolescents report using to regulate their emotions and perceptions of QOL across 10 dimensions. Specifically, Table 4 shows that more frequent use of internal-dysfunctional and external-dysfunctional emotion regulation was significantly associated with lower perceived QOL for most dimensions. In contrast, the use of internal-functional and external-functional emotion regulation was significantly associated with increased QOL in most dimen-

Table 4. Correlations between emotion regulation (ER) scales (Regulation of Emotions Questionnaire) and psychosomatic difficulties scale

	Internal-dysfunctional ER	Internal-functional ER	External-dysfunctional ER	External-functional ER
Psychosomatic difficulties	0.58*	NS	0.29*	NS

*Correlation is significant at the 0.01 level.

NS = not significant.

Table 5. Correlations between the four emotion regulation (ER) scales (Regulation of Emotions Questionnaire) and the 10 Quality of Life (QOL) (KIDSCREEN) scales

	Internal-dysfunctional ER	Internal-functional ER	External-dysfunctional ER	External-functional ER
QOL: physical well-being	-0.26**	0.18**	-0.07	0.33**
QOL: psychological well-being	-0.42**	0.28**	-0.14*	0.41**
QOL: moods and Emotions	-0.59**	0.16*	-0.22**	0.18**
QOL: social support	-0.24**	0.28**	-0.16*	0.46**
QOL: parents/home life	-0.30**	0.17*	-0.26**	0.32**
QOL: self-perception	-0.46**	0.14*	-0.17*	0.18*
QOL: autonomy	-0.24**	0.18**	-0.12	0.32**
QOL: school environment	-0.28**	0.30**	-0.19**	0.45**
QOL: bullying	-0.32**	0.04	-0.34**	0.11
QOL: financial resources	-0.18**	0.20**	-0.12	0.32**
Total QOL	-0.47**	0.31**	-0.27**	0.46**

*Correlation is significant at the 0.05 level. **Correlation is significant at the 0.01 level.

sions. Also, since there were differences in relationships with QOL for internal and external scales, this shows that internal and external scales tap into something different and should not be combined in the measure since important information would be lost.

DISCUSSION

A new measure of individual differences in emotion regulation (the REQ) was developed and tested in a sample of adolescents. The measure was based on literature which suggested that strategies used to regulate emotions may include those that facilitate processing of emotions (adaptive or functional strategies) and those that do not (maladaptive or dysfunctional). The literature also pointed to strategies that regulate emotions through drawing on internal (internalized or intrapersonal) resources or external (actual attachment or interpersonal) resources. Therefore, the measure incorporated scales to assess internal-dysfunctional, internal-functional, external-dysfunctional and external-functional emotion regulation strategies. The data collected from a sample of 225 adolescents supported the structure of the measure and

its scales in factor analyses. Also, the internal consistency of each scale was good, with the exception of the external-functional scale, which had a Cronbach's alpha of below the accepted level of 0.7.

The item assessing the use of exercise as an external-functional emotion regulation strategy fits less well with its scale than with the other items, thus reducing the internal consistency of the scale. This may be because it is a rather ambiguous item for an emotion regulation strategy and can actually have both functional and dysfunctional implications depending on the precise nature of its use. For example, exercise can successfully reduce negative emotion (e.g., Thayer, Newman, & McClain, 1994) that an individual may not attempt to inhibit, but it can also become a compulsive addiction (e.g., Christo et al., 2003). The latter use is clearly more suited to inclusion on the external-dysfunctional scale and can be linked to inhibition of emotions. Future efforts to improve this particular scale, perhaps with the addition of extra relevant items, are planned.

There is a strong association between the two forms of dysfunctional emotion regulation, as well as between the two forms of functional emotion regulation. However, all four scales show different relationships with the other variables measured;

these will be discussed in detail below. The internal- versus external modes of regulation capture additional information and, therefore, a model of dysfunctional versus functional emotion regulation would be too simplistic.

Relationships with Other Measures

Expected relationships were identified between the scales on the new emotion regulation measure and (1) emotional and behavioural problems, (2) psychosomatic health problems and (3) QOL, thus providing evidence for the validity of the measure. The use of both internal-dysfunctional and external-dysfunctional emotion regulation is related to emotional and behavioural problems, assessed with the SDQ (Goodman, 1997). Adolescents who reported the frequent use of dysfunctional strategies to regulate emotions exhibited more emotional and behavioural problems according to their parents. Indeed, the frequent use of both types of dysfunctional emotion regulation was related to increased severity of emotional symptoms, conduct problems, hyperactivity and peer problems. This finding supports the suggestion made by Gross (1999) that dysfunctional emotion regulation is implicated in many of the psychiatric disorders classified by DSM-IV (American Psychiatric Association, 1994).

The strongest relationships between dysfunctional emotion regulation and specific problems are those between internal-dysfunctional emotion regulation and emotional symptoms, an 'internalizing' problem (e.g., Casey, 1996), and between external-dysfunctional emotion regulation and conduct problems, an 'externalizing problem'. This supports the views of Caspi, Henry, McGee, Moffit, and Silva (1995) and Southam-Gerow and Kendall (2002), who argue that dysfunctional emotion regulation mirrors the traditional classification of psychological disorders into internalizing and externalizing problems. It also further supports the validity of this new measure.

The frequent use of internal-functional emotion regulation strategies was associated with less emotional and behavioural problems overall but showed no significant negative relationship to any specific disorders. Adolescents who reported frequently using external-functional emotion regulation strategies exhibited fewer peer problems and greater pro-social behaviour according to their parents. This can be explained by the fact that the majority of items in the external-functional scale

relate to seeking social support, a commonly used method of emotion regulation. 'Unsuccessful' emotion regulation predicts poor peer relationships (Eisenberg & Fabes, 1992); it is also evident that 'successful' (or functional) emotion regulation is associated with fewer peer problems and greater pro-social behaviour, as perceived by parents.

Adolescents who reported the frequent use of internal-dysfunctional and external-dysfunctional strategies were more likely to report the experience of psychosomatic health problems. Since dysfunctional emotion regulation is defined as the use of strategies that reject or block emotional experience, this finding is consistent with research that shows a relationship between emotional inhibition and illness (e.g., Pennebaker, 1990). Furthermore, the fact that reported use of both internal-dysfunctional and external-dysfunctional emotion regulation was associated with psychosomatic health problems supports Siegeman's (1993) proposal that a focus on either maladaptive inhibitory or expressive strategies is linked to poor physical health, thus providing further support for the validity of the new measure.

Reported use of the four emotion regulation strategies is significantly associated with perceptions of QOL. Indeed, frequent use of both internal-functional and external-functional strategies is associated with a better overall QOL. In contrast, frequent use of both internal-dysfunctional and external-dysfunctional strategies is associated with a reduced overall QOL. The KIDSCREEN QOL instrument (Ravens-Sieberer et al., 2001, 2005) used in this research is a generic QOL measure, designed to assess childhood functioning (e.g., Spieth, 2001). Therefore, the significant association observed between dysfunctional emotion regulation and reduced QOL supports the claim by Kostiuik and Fouts (2002) that unsuccessful emotion regulation impairs overall functioning.

Importantly, functional emotion regulation is associated with well-being across many QOL dimensions. Indeed, with the exception of the QOL dimension assessing bullying, the frequent use of both internal-functional and external-functional emotion regulation strategies is associated with better QOL in every dimension. Functional emotion regulation is often not assessed with other measures. However, the current research shows that assessing this aspect is important, since it is related to aspects of healthy functioning in terms of perceived QOL.

There are a number of limitations with this study that should be taken into consideration. First,

adolescents were not included in the generation of items for the measure. However, the authors drew on their clinical experience with adolescents, as well as that of the other professionals consulted as part of the measure construction. Another limitation is the use of the same sample to conduct both exploratory and confirmatory factor analyses. Unfortunately the sample was not large enough to use the split-half method. Although the structure of the measure was supported, these analyses should be viewed as exploratory. Even though the measure showed some promising results, it is important to remember that there are difficulties inherent in measuring emotion regulation by self-report. For example, it is not possible for people to accurately report on all aspects of their emotion regulation, since many strategies may be unconscious (e.g., Gross, 1998). Also, young people may describe very different emotion regulation strategies to those that they actually use in emotion-provoking situations. Despite these difficulties, tools that assess individual differences in emotion regulation are needed and may be used alongside other methods of assessing emotion regulation in order to gain a comprehensive picture.

In summary, a new measure of individual differences in adolescent emotion regulation was developed through drawing on expert opinion. The new measure showed good psychometric properties when used with a population of more than 200 adolescents. Indeed, the data supported the structure of the scale, which assesses four types of emotion regulation: internal-dysfunctional, internal-functional, external-dysfunctional and external-functional. The measure showed expected relationships with measures of emotional and behavioural problems, psychosomatic health problems and QOL. These results provide support for the validity of the measure, confirming the important role that emotion regulation plays in different aspects of health and functioning in young people. This measure of individual differences in emotion regulation in young people may therefore prove a useful tool for future research.

REFERENCES

- American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington DC: American Psychiatric Association.
- Bentler, P.M. (1989). *EQS structural equations program manual*. Encino, CA: Multivariate Software.
- Casey, R.J. (1996). Emotional competence in children with externalizing and internalizing disorders. In M. Lewis, & M.W. Sullivan (Eds), *Emotional development in atypical children* (pp. 161–183). Mahwah, NJ: Lawrence Erlbaum Associates.
- Caspi, A., Henry, B., McGee, R.O., Moffit, T.E., & Silva, P.A. (1995). Temperamental origins of child and adolescent behaviour problems: From age three to age fifteen. *Child Development*, 66, 55–68.
- Chapman, A.L., Specht, M.W., & Celluci, T. (2005). Borderline personality disorder and deliberate self-harm: Does experiential avoidance play a role? *Suicide and Life Threatening Behavior*, 35(4), 388–399.
- Christo, G., Jones, S.L., Haylett, S., Stephenson, G.M., Lefever, R.M., & Lefever, R. (2003). The Shorter PROMIS Questionnaire: Further validation of a tool for simultaneous assessment of multiple addictive behaviours. *Addictive Behaviors*, 28(2), 225–248.
- Comrey, A.L., & Lee, H.B. (1992). *A first course in factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Currie, C., Samdal, O., Boyce, W., & Smith, R. (Eds). (2001). *Health behaviour in school-aged children: A WHO cross-national study (HBSC)*. Research Protocol for the 2001/2002 Survey. Child and Adolescent Health Research Unit (CAHRU), University of Edinburgh.
- Eisenberg, N., & Fabes, R.A. (1992). *Emotion and its regulation in early development*. San Francisco, CA: Jossey-Bass.
- Eisenberg, N., Champion, C., & Ma, Y. (2004). Emotion-related regulation: An emerging construct. *Merrill-Palmer Quarterly*, 50.3, 236–259.
- Garnefski, N., Kraaij, V., & Spinhoven, P. (2001). Negative life events, cognitive emotion regulation and emotional problems. *Personality and Individual Differences*, 30, 1311–1327.
- Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology, Psychiatry, and Allied Disciplines*, 40(5), 791–799.
- Gratz, K.L., & Roemer, L. (2004). Multidimensional assessment of emotion regulation and dysregulation: Development, factor structure, and initial validation of the Difficulties in Emotion Regulation Scale. *Journal of Psychopathology and Behavioral Assessment*, 26, 41–54.
- Gross, J.J. (1998). The emerging field of emotion regulation: An integrative review. *Review of General Psychology*, 2(3), 271–299.
- Gross, J.J. (1999). Emotion regulation: Past, present, future. *Cognition and Emotion*, 13(5), 551–573.
- Gross, J.J., & John, O.P. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology*, 85, 348–362.
- Gross, J.J., & Munoz, R.F. (1995). Emotion regulation and mental health. *Clinical Psychology: Science and Practice*, 2, 151–164.
- John, O.P., & Gross, J.J. (2004). Healthy and unhealthy emotion regulation: Personality processes, individual differences, and lifespan development. *Journal of Personality*, 72, 1301–1334.
- Kostiuk, L.M., & Fouts, G.T. (2002). Understanding of emotions and emotion regulation in adolescent

- females with conduct problems: A qualitative analysis. *The Qualitative Report*, 7(1), 1–10.
- Lazarus, R.S., & Folkman, S. (1984). *Stress, appraisal, and coping*. New York: Springer.
- Linehan, M. (1993). *Cognitive-behavioral treatment of borderline personality disorder*. London: Guilford.
- Losoya, S., Eisenberg, N., & Fabes, R.A. (1998). Developmental issues in the study of coping. *International Journal of Behavioural Development*, 22(2), 287–313.
- Pennebaker, J.W. (1990). *Opening up: The healing powers of confiding in others*. New York: Morrow.
- Power, M.J., & Dalgleish, T. (1997). *Cognition and emotion: From order to disorder*. Hove: Psychology Press.
- Ravens-Sieberer, U., Gosch, A., Abel, T., Aucquier, P., Bellach, B., Bruil, J., Dur, W., Power, M., Rajmil, L., & the European KIDSCREEN Group. (2001). Quality of life in children and adolescents: A European public health perspective. *Social and Preventative Medicine*, 46, 294–302.
- Ravens-Sieberer, U., Gosch, A., Rajmil, L., Erhart, M., Bruil, J., Duer, W., Auquier, P., Power, M., Abel, T., Czemy, L., Mazur, J., Czimbalmos, A., Tountas, Y., Hagquist, C., Kilroe, J., & the European KIDSCREEN Group. (2005). KIDSCREEN-52 quality-of-life measure for children and adolescents. *Expert Review of Pharmacoeconomics & Outcomes Research*, 5(3), 353–364.
- Scheier, M.F., Weintraub, J.K., & Carver, C.S. (1986). Coping with stress: Divergent strategies of optimists and pessimists. *Journal of Personality and Social Psychology*, 51, 1257–1264.
- Shields, A., & Cicchetti, D. (1997). Emotion regulation among school-age children: The development and validation of a new criterion Q-sort scale. *Developmental Psychology*, 33, 906–916.
- Siegeman, A.W. (1993). Cardiovascular consequences of expressing, experiencing and repressing anger. *Journal of Behavioural Medicine*, 16, 539–569.
- Southam-Gerow, M.A., & Kendall, P.C. (2002). Emotion regulation and understanding. Implications for child psychopathology and therapy. *Clinical Psychology Review*, 22, 189–222.
- Spieth, L.E. (2001). Generic health-related quality of life measures for children and adolescents. In H.M. Koot, & J.L. Wallander (Eds). *Quality of life in children and adolescents: Concepts, methods and findings*. Hove: Brunner-Routledge.
- Tabachnick, B.G., & Fidell, L.S. (2001). *Using multivariate statistics* (4th ed.). New York: HarperCollins.
- Tabachnick, B.G., & Fidell, L.S. (2001). *Using multivariate statistics* (4th ed.). New York: HarperCollins.
- Thayer, R.W., Newman, I.R., & McClain, T.M. (1994). Self-regulation of mood: Strategies for changing a bad mood, raising energy, and reducing tension. *Journal of Personality and Social Psychology*, 67, 910–925.
- Thompson, R.A. (1994). Emotion regulation: A theme in search of definition. In N.A. Fox (Ed.), *The development of emotion regulation: Biological and behavioural considerations*. *Monograph of the Society for Research in Child Development*, 59, 25–52.
- Weiner, B. (1986). *An attributional theory of motivation and emotion*. New York: Springer-Verlag.

Copyright of *Clinical Psychology & Psychotherapy* is the property of John Wiley & Sons Ltd. 1996 and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.

Appendix B: Sample Scales and Instruments

Zhuang, X., MacCann, C., Wang, L., Liu, L., & Roberts, R.D. (2008). *Development and Validity Evidence Supporting a Teamwork and Collaboration Assessment for High School Students*. (p.18). ETS RR-08-50. Educational Testing Service: Princeton, NJ.

Table 1***Factor Loadings of the Revised Student Self-Report Teamwork Scale***

Items	F1	F2	F3	Mean (SD)
I enjoy bringing team members together	.75			4.45 (1.29)
I share ideas with others to accomplish a task	.74			4.58 (1.20)
I acknowledge the accomplishments of my peers	.71			4.71 (1.19)
I enjoy helping team members	.67			4.77 (1.19)
I value different perspectives to help me strengthen my understandings of issues or problems	.67			4.40 (1.26)
I provide appropriate feedback to team members	.60			4.16 (1.18)
I think that exchange of ideas among team members can lead to creative solutions	.60			4.84 (1.26)
I cooperate with other students	.53			4.72 (1.12)
I like team activities	.46			4.43 (1.33)
I am inspired by others' ideas and thoughts	.46			4.09 (1.25)
I contribute to the definition of a team's goals	.43			4.50 (1.22)
I respect the opinions of my peers	.41		.34	4.76 (1.11)
I like to be in charge of groups or projects		.70		3.75 (1.42)
I know how to make other students see things my way		.65		4.03 (1.12)
I can convince my peers about anything		.57		3.80 (1.19)
I believe I am a good leader		.56		4.68 (1.36)
I am comfortable providing constructive criticism		.44		3.80 (1.46)
I carefully consider the facts to persuade my peers		.43		4.30 (1.29)
I seek to influence my peers		.41		3.61 (1.44)
I suggest alternative solutions to problems		.30		3.99 (1.23)
My arguments are constructive		.23		4.15 (1.27)
I am a good listener			.72	4.91 (1.11)
I am open to varying opinions			.65	4.64 (1.23)
I take other students' interests into account			.55	4.32 (1.28)
I adapt to change well			.53	4.16 (1.15)
I am flexible in team situations			.51	4.68 (1.16)
I believe that there is only one "best" solution			.49	4.38 (1.37)
I dislike it when people challenge my views ^a			.38	3.92 (1.29)
I understand that each team member is different	.30		.31	5.33 (1.04)
I consider team members first			.29	3.98 (1.36)

Note. F1 = cooperation; F2 = advocating/influence; F3 = negotiation.

^a Reverse-scored item.

Appendix B: Sample Scales and Instruments

Zimet, G.D., Dahlem, N.W., Zimet, S.G. & Farley, G.K. (1988). The Multidimensional Scale of Perceived Social Support. *Journal of Personality Assessment*, 52, 30-41.

Multidimensional Scale of Perceived Social Support (Zimet, Dahlem, Zimet & Farley, 1988)

Instructions: We are interested in how you feel about the following statements. Read each statement carefully. Indicate how you feel about each statement.

Circle the "1" if you **Very Strongly Disagree**
 Circle the "2" if you **Strongly Disagree**
 Circle the "3" if you **Mildly Disagree**
 Circle the "4" if you are **Neutral**
 Circle the "5" if you **Mildly Agree**
 Circle the "6" if you **Strongly Agree**
 Circle the "7" if you **Very Strongly Agree**

1.	There is a special person who is around when I am in need.	1	2	3	4	5	6	7	SO
2.	There is a special person with whom I can share my joys and sorrows.	1	2	3	4	5	6	7	SO
3.	My family really tries to help me.	1	2	3	4	5	6	7	Fam
4.	I get the emotional help and support I need from my family.	1	2	3	4	5	6	7	Fam
5.	I have a special person who is a real source of comfort to me.	1	2	3	4	5	6	7	SO
6.	My friends really try to help me.	1	2	3	4	5	6	7	Fri
7.	I can count on my friends when things go wrong.	1	2	3	4	5	6	7	Fri
8.	I can talk about my problems with my family.	1	2	3	4	5	6	7	Fam
9.	I have friends with whom I can share my joys and sorrows.	1	2	3	4	5	6	7	Fri
10.	There is a special person in my life who cares about my feelings.	1	2	3	4	5	6	7	SO
11.	My family is willing to help me make decisions.	1	2	3	4	5	6	7	Fam
12.	I can talk about my problems with my friends.	1	2	3	4	5	6	7	Fri

The items tended to divide into factor groups relating to the source of the social support, namely family (Fam), friends (Fri) or significant other (SO).

Appendix B: Sample Scales and Instruments

Freien Universität Berlin (n.d.). *The General Self-Efficacy Scale (GSE)*. Retrieved October 5, 2009, from <http://userpage.fu-berlin.de/~health/engscal.htm>

Appendix	English version by Ralf Schwarzer & Matthias Jerusalem, 1993	
	1	I can always manage to solve difficult problems if I try hard enough.
	2	If someone opposes me, I can find the means and ways to get what I want.
	3	It is easy for me to stick to my aims and accomplish my goals.
	4	I am confident that I could deal efficiently with unexpected events.
	5	Thanks to my resourcefulness, I know how to handle unforeseen situations.
	6	I can solve most problems if I invest the necessary effort.
	7	I can remain calm when facing difficulties because I can rely on my coping abilities.
	8	When I am confronted with a problem, I can usually find several solutions.
	9	If I am in trouble, I can usually think of a solution.
	10	I can usually handle whatever comes my way.
Response Format	1 = Not at all true 2 = Hardly true 3 = Moderately true 4 = Exactly true	

Appendix B: Sample Scales and Instruments

Schwarzer, R., Diehl, M., & Schmitz, G.S. (1999). *Self-Regulation Scale*. Retrieved October 5, 2009, from http://userpage.fu-berlin.de/~health/selfreg_e.htm

Self-Regulation

[Ralf Schwarzer](#), Manfred Diehl, & Gerdamarie S. Schmitz, 1999

This scale refers to post-intentional self-regulation when individuals are in the phase of goal-pursuit and face difficulties in maintaining their action. In such a **maintenance situation** it is required to focus attention on the task at hand and to keep a favorable emotional balance. Thus, **attention-regulation** and **emotion-regulation** are reflected in these scale items.

In a sample of N = 442 persons the scale has obtained an internal consistency of **Cronbach's alpha** = .76. In a sample of N = 239 persons the scale yielded a **retest stability** of .62 after six weeks.

There were **associations** found with [general self-efficacy](#) beliefs ($r = .57$), and with [proactive coping](#) ($r = .55$).

1. I can concentrate on one activity for a long time, if necessary.
2. If I am distracted from an activity, I don't have any problem coming back to the topic quickly.
3. If an activity arouses my feelings too much, I can calm myself down so that I can continue with the activity soon.
4. If an activity requires a problem-oriented attitude, I can control my feelings.
5. It is difficult for me to suppress thoughts that interfere with what I need to do. (–)
6. I can control my thoughts from distracting me from the task at hand.
7. When I worry about something, I cannot concentrate on an activity. (–)
8. After an interruption, I don't have any problem resuming my concentrated style of working.
9. I have a whole bunch of thoughts and feelings that interfere with my ability to work in a focused way. (–)
10. I stay focused on my goal and don't allow anything to distract me from my plan of action.

Note: (–) indicates the item has to be reversed.

Response format:

(1) not at all true, (2) barely true, (3) moderately true, (4) exactly true