Self-Evaluations and Performance: Evidence from Adolescence.

Oliver Himmler[a]
MPI Bonn

Tobias Koenig[b]
Hannover University

Leibniz University Hannover, Discussion Paper No. 502

SSN 0949-9962

October 2012

(first version: March 2011)

Abstract

A positive view of the self is often portrayed as a valuable asset in the sense that it can have performance enhancing properties. Using data on self-esteem – the most fundamental manifestation of positive self evaluations – and high school grade point averages of American students we produce results in line with this idea and find a positive link between favorable self-evaluations and higher levels of educational performance. However, when we exploit exogenous variation in self-esteem due to adolescent skin problems in order to account for the possible endogeneity of self-esteem, this finding is reversed and we obtain a negative effect on performance. We discuss mechanisms that may generate such an adverse causal effect of positive self-evaluations, and conclude that self-esteem and effort need not always be complements but can actually be substitutes.

JEL Codes: I2, J24.

Keywords: self-evaluations, self-esteem, non-cognitive skills, human capital, performance.

[a]Max Planck Institute for Research on Collective Goods, Kurt-Schumacher-Strasse 10, 53113 Bonn, Germany. email: himmler@coll.mpg.de

[b]Hannover University, Department of Economics, Koenigsworther Platz 1, 30167 Hannover, Germany. email: koenig@sopo.uni-hannover.de

We thank conference and seminar participants at UC San Diego, Maastricht University, Max-Planck-Institute Bonn, University of Zurich, NHH Bergen, Mannheim, Trier, Goettingen, the 68th Congress of the International Institute of Public Finance in Dresden, the 24th Annual Conference of the European Association of Labor Economists in Bonn and the 2012 Congress of the European Economic Association in Malaga for their very helpful comments and suggestions.

1 Introduction

A positive evaluation of the self is often seen as crucial for human motivation and performance. From ancient Greek philosophy to the rise of modern psychology in the 19th century, scholars have suggested that appreciating oneself is a key prerequisite for personal development and for conducting a successful life (Duckworth et al., 2005). Today this view that self-esteem is an important asset has become ingrained in everyday western culture and public opinion (Putnam, 2001). Hundreds of books and magazines praise a positive self view as the magic bullet for navigating the challenges in life such as child rearing, dieting, mating, or maintaining a successful professional career (Salerno, 2005). American schools and the juvenile justice system devote large resources to self-esteem building programs with the goal of fostering a positive self image in adolescents and thereby combating teenage ills like violence and delinquency, unwanted pregnancy and high school dropout. Politicians have sometimes even claimed that individual self-esteem is essential for a nation's prosperity and economic growth (Baumeister et al., 2003). Against this backdrop, our paper is the first to exploit a quasi-natural experiment in order to provide causal evidence on the effects that positive self-evaluations may have on performance. Specifically, by exploiting exogenous variation in self-image due to skin problems in adolescence we evaluate whether self-esteem influences academic performance in school.

Self-esteem is considered to be the most fundamental manifestation of self evaluations (Judge et al. 1997, Judge and Bono 2001) and the concept is widely used in other social sciences. Yet it is only with the recent emergence of a fast growing literature on the economic effects of non-cognitive skills that the concept has gained the attention of economists (see Bowles et al., 2001; Borghans et al., 2008; and Almlund et al., 2011 for surveys). The theoretical idea of why positive self-views can improve performance – and in that sense have instrumental value – is that they serve as complements to effort. Every task or project usually comes with a positive probability of failure, and a bigger ego may help to protect against psychological or physical distress that arises from the fear of performing badly. This anxiety-buffering function of positive self-attributions can not only

increase initial effort or investments. It also is assumed to help suppress negative emotions once a task has been started and setbacks occur. This may positively contribute to effort by nourishing economically valuable features such as perseverance and persistence (Benabou and Tirole, 2002; Compte and Postlewaite, 2004).[1] A considerable body of empirical literature has produced evidence in line with this reasoning and the popular credo that positive self-views are conducive to success: they have proved to be powerful predictors of higher wages, better educational outcomes, good health and less involvement in crime – sometimes even more so than measures of cognitive ability such as IQ and achievement test scores (Heckman et al., 2006).

Empirically relating favorable views of the self to performance raises methodological issues. Is it the psyche that determines economic outcomes, or do economic outcomes determine the psyche? Obviously this is a metaphysical question, yet at least in the context of positive self-views it is conceivable that greater economic success invokes more favorable assessments of the self. Failure to take this possible backward causation into account will give rise to upward-biased estimates. Similarly, variables such as cognitive ability and family background are likely to be positively linked to both to a better self-image and increased performance; unfortunately they are often also unobserved by the researcher and will then also generate exaggerated coefficients. An important step in the literature has been to address these endogeneity issues by using lagged measures of the psychological variable of interest (Feinstein, 2000; Mueller and Plug, 2006; Wadell, 2006; Fortin, 2008; Drago, 2011). This idea relies on the assumption that in such a model backwards causality can be ruled out because personality measured before the outcome variable cannot be affected by present outcomes. As Almlund et al. (2011) point out, a complicating factor is that individuals might anticipate their future success and thus have a more favorable view of the self today, in which case the upward bias will remain. In the event that unobservable characteristics driving personality and outcomes are correlated over time, the use of lagged variables may still face unobserved heterogeneity issues. In that case the same omitted

[1] While psychologists refer to this mechanism as the anxiety-buffering function of self-esteem (Greenberg et al., 1992), Compte and Postlewaite (2004) call this confidence-enhanced performance. Some authors even go so far as to suggest self-esteem as a new form of capital which they call 'psychological capital' (see, e.g., Goldsmith et al., 1997).

variables as in regressions without the lagged personality variable can induce bias.[2]

We propose a different identification strategy to address these issues: This paper is the first to provide causal evidence on the effects of a positive self-image on performance from a quasi-natural experiment. Our measure of global self-evaluations is the Rosenberg self-esteem scale, which inquires about feelings of self-worth but also encompasses an evaluation of how an individual feels about their qualities and how they generally rate their performance in life. Because our aim is to use exogenous variation in self-esteem and because these self-evaluations tend to be more stable in adults, our search for an instrument focuses on a adolescence, a period in life that is crucial for the formation of self-esteem. In this period, some teenagers are exposed to a twist of fate and develop skin problems in the form of acne while others are more lucky and do not. We use this variation in skin conditions as an instrument for self-esteem and analyze effects on performance in one of the most important areas in an adolescent's life: school. The extent to which an individual performs well in school is measured by grades we obtain from high school transcripts.

In the course of the paper we will explain in much more detail, but the basic idea of why in this setup acne is a valid instrument goes as follows: Acne produces significant drops in adolescent self-esteem, yet even though it can have severe effects on an individual's psyche, acne is primarily a cosmetic problem and it does not do any harm to hard skills such as cognitive ability. Because it does not cause any physical impairment that might directly interfere with school performance either, the argument is that acne, conditional on certain covariates, will not affect grades other than through self-esteem.

Because acne is inextricably linked to maturity levels, and anecdotally an outcome of lifestyle choices such as diet and hygiene, we lay out the conditions under which acne can be considered to be random. We provide a thorough survey of the medical literature which shows that developing acne in puberty is indeed not an outcome of individual choices. Notably, there is no evidence that acne develops from individual behavior like poor hygiene or

[2]Another approach is to regress early life personality on the respondent's age at the time the psychological data were collected. The residuals are then taken as predictors meant to measure a person's age-corrected personality (see, e.g., Heineck and Anger, 2010; Nyhus and Pons, 2005; Osborne Groves, 2005). If age is the only determinant of personality this strategy is sufficient. If other factors also determine personality then the correlation with the error is mitigated but may not be fully canceled out.

consuming unhealthy foods. Holding constant some key factors such as the stage of physical development, the availability of medical treatment, and the use of oral contraceptives as well as controlling for possible discrimination by others, we argue that acne generates quasi-experimental variation in self-esteem which can be used to identify the effects of self-views on school performance.

When using standard methods that do not account for endogeneity, we find that there is a positive correlation between high self-evaluations and educational outcomes. Our instrumental variable results confirm that there is indeed a causal effect of self-esteem on school performance, and in that sense the results support the hypothesis that psychological factors matter. However, the effect goes in the opposite direction of what is usually postulated and the pure correlations that we have found. Higher levels of self-esteem in adolescence lead to *worse* academic performance, and the estimated effect is substantial. This negative relationship between self-esteem and school performance holds across a variety of models: Lagged dependent variable regressions account for the possibility that having particularly good or bad grades in school may promote acne via increased levels of stress. A first differenced approach rules out that our results are driven by unobserved fixed factors that foster good school performance and might at the same time be related to skin problems. In addition, we provide an array of falsification tests which show that the effect we measure is not simply driven by a confoundedness of acne with early maturity or heterogeneity in hormone levels across individuals.

We discuss and provide empirical support for mechanisms that can plausibly generate such an adverse effect of self-esteem on task performance. Our results indicate that those students with high levels of self-esteem have a higher probability of working in a side job, and are more socially active. One possible explanation for this is that individuals with high global self-esteem are also more confident in reaching their academic goals, which under reference point preferences would give them reason to reduce effort in school and turn to other tasks. An alternative explanation is that self-esteem induces changes in preferences themselves: As has recently been pointed out by psychologists, holding positive self-views can foster experimentation and therefore interest in non-academic activities (see,

e.g., Baumeister, 2003). Our results are in line with both these explanations and contribute a new angle to the literature by providing first evidence that positive self-views and effort need not always be complements, but can actually be *substitutes* – a possibility that has received little attention in the recent economics literature on non-cognitive skills.

The remainder of the paper is organized as follows. Section 2 describes the data, gives some descriptive statistics and presents (conditional) correlations between self-esteem and high school grades. Section 3 explains our identification strategy in detail. Section 4 presents the baseline IV results and lays out robustness checks and falsification tests. Channels are presented in section 5. Section 6 concludes.

2 Data and correlation between self-esteem and grades

Before we introduce the instrumental variable setup in chapter 3, this section of the paper provides a description of the data we use, some evidence on the correlational link between self-esteem and grades, as well as arguments as to why these correlations need not identify a causal channel from personality features to performance.

Our data stem from the restricted version of the AddHealth survey, which followed a nationally representative sample of young Americans over a period of 16 years from early adolescence to adulthood in four waves. The first wave was administered during the 1994/95 school year, and more than 20,000 students in more than one hundred schools across the United States participated in the survey. Roughly one year later, Wave II re-interviewed the approximately 15,000 students that were still in school at the time. Since we focus on school outcomes and by Wave III the respondents were already out of school, this paper uses data from the first two waves. The main reason for choosing the AddHealth survey is that it provides an unusually rich set of variables that describe the lives of adolescents. Among them are a measure of self-esteem as well as a question on skin problems which provides the exogenous variation that our identification approach will rely on.

The quantification of self-esteem has a long standing tradition in the field of psychology,

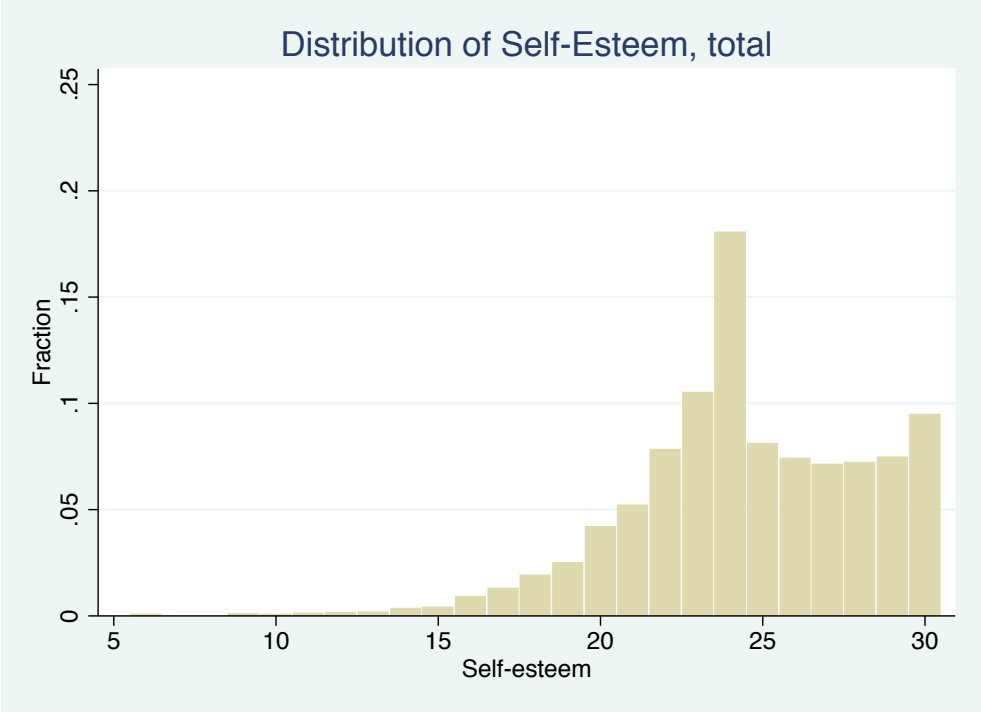
going back to at least the 1960s. Rosenberg (1965) introduced what was to become the workhorse of self-esteem research in psychology and economics (Almlund et al., 2011) up to this day: the Rosenberg self-esteem scale. The original Rosenberg scale measures evaluations of self-worth and capabilities by asking a battery of ten questions, each of which has four answer categories: "strongly disagree","disagree", "agree", and "strongly agree". The AddHealth data provide a shortened version of the original Rosenberg scale: it asked six of the original questions while adding a fifth neutral answer category. Specifically, the six questions read: "Do you agree or disagree that you..." (1) "[...] have many good qualities", (2) "[...] have a lot to be proud of", (3) "[...] like yourself just the way you are", (4) "[...] feel you are doing things just about right", (5) "[...] feel socially accepted", (6) "[...] feel loved and wanted". We reverse the coding from AddHealth such that "strongly agree" scores 5 points, and "strongly disagree" scores 1 point. Aggregating the scores from all items, we obtain the AddHealth pendant of the Rosenberg scale with a minimum score of 6, and a maximum score of 30.

The top left corner of Figure 1 shows the distribution of values scored on the Rosenberg scale for Wave I.[3] Two things stand out about this graph: most individuals score rather high on the scale, but only less than 10 percent of respondents choose the highest category in all questions. The modal score is 24, which corresponds to picking – on average – the second highest category on all questions. Values below fifteen are very rare, less than one percent of the estimation sample score in this range.

As a measure for students' performance in school, we take the grade point average (GPA) of a respondent in the year of the survey across all subjects. GPA data comes from high school transcripts that AddHealth has obtained from the participating schools – a fact that renders them an impartial measure of educational achievement. On the downside, AddHealth could not obtain transcripts for all respondents and this significantly reduces our estimation sample. The average GPA is roughly 2.6, where a value of 0 corresponds to failing all classes and 4 means having attained all "A"s. Figure 2 plots the average GPA

[3]Figures, descriptives and estimations in this chapter are shown for Wave I data and use the observations that are in our baseline instrumental variable sample used in Table 2.

Figure 1: Distribution of self-esteem, Wave I.



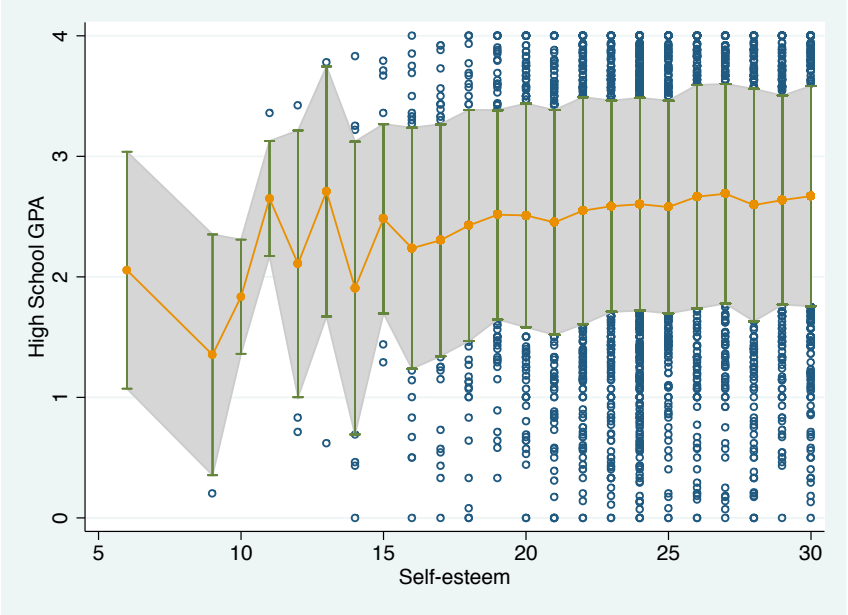
Note: The figure shows the distribution of self-esteem. the sample is the same as the estimation sample used in Table 2. Number of observations $N = 4927$.

across self-esteem levels. Each circle represents one observation, and the area shaded in gray marks a standard deviation above and below the mean GPA for each self-esteem level. Evidently, the higher an individual's level of self-esteem, the higher is his attained grade point average.[4] This descriptive observation is very much in line with what is typically found in the literature and with the arguments brought forward by proponents of self-esteem boosting programs. If interpreted causally, this indeed suggests that there may be benefits to holding positive self-views in terms of attaining better grades.

The positive correlation is confirmed when looking at Table 1, where we summarize the baseline OLS estimates. Column (1) shows the pure correlation between self-esteem and grades, without adding any controls – basically, this reproduces the relationship from Figure 2. Some basic individual characteristics are added in column (2): age fixed effects, race, gender and health status. All of these controls are highly statistically significant, yet the association between self-esteem and GPA remains stable. Because it seems plausible that innate or acquired ability as well as family background would affect both self-views and

[4]Disregarding the kinks in the low range of self-esteem, where the Rosenberg index is below fifteen. As stated above and as can be seen from the graph, these categories are very sparsely populated and thus do not contradict the statement.

Figure 2: Self-esteem and high school GPA, Wave I.



Note: The figure displays the mean GPA by self-esteem levels, as measured by the Rosenberg index. The shaded area corresponds one standard deviation. The sample is the same as the estimation sample used in Table 2. Number of observations $N = 4927$.

school outcomes, column (3) adds the score from the Peabody Picture Vocabulary Test which measures verbal ability, a measure of household income, whether the respondent lives with a single parent, whether the parents are on welfare and the highest level of education obtained by either parent. All additional controls have the expected sign and are mostly also highly significant. Column (4) adds grade fixed effects as well as their interaction with age since how old a respondent is in comparison to his peers might matter both for the respondent's self-esteem as well as their grades, e.g. due to relative grading practices. Finally, school fixed effects are held constant in column (5). The upshot of these regressions is that the positive correlation between Rosenberg self-esteem and school grades seems to be quite robust when a fairly standard set of individual and school characteristics is taken into account. In terms of magnitude however, this association is not overly strong: a coefficient of around 0.02 means that a one standard deviation change in self-esteem is linked to a roughly 0.07 increase in GPA, which corresponds to 7.5 percent of a standard deviation.[5]

[5]While this may not be a small effect for a policy intervention, the effects of non-cognitive skills such as self esteem should rather be compared to the effects of cognitive skills – such as the Peabody test score, where in this estimation a one standard deviation increase is linked to an increase of 30 percent of a standard deviation in GPA.

Table 1: OLS, dependent variable high school GPA, Wave I.

	(1)	(2)	(3)	(4)	(5)
self-esteem	.0195*** (0.004)	.0197*** (0.004)	.0198*** (0.004)	.0193*** (0.003)	.0176*** (0.003)
basic individual characteristics					
male		−.365*** (0.018)	−.396*** (0.022)	−.376*** (0.021)	−.366*** (0.019)
health very good[(a)]		−.052** (0.023)	−.0726*** (0.022)	−.0757*** (0.022)	−.0785*** (0.022)
health good		−.298*** (0.035)	−.24*** (0.032)	−.233*** (0.031)	−.239*** (0.031)
health fair		−.499*** (0.056)	−.395*** (0.057)	−.369*** (0.057)	−.357*** (0.052)
health poor		−.506*** (0.120)	−.439*** (0.164)	−.471*** (0.160)	−.486*** (0.159)
black[(b)]		−.418*** (0.062)	−.195*** (0.055)	−.193*** (0.056)	−.15*** (0.047)
am. indian		−.315*** (0.058)	−.165 (0.106)	−.159 (0.109)	−.154 (0.128)
asian		.255*** (0.083)	.338*** (0.045)	.322*** (0.044)	.384*** (0.041)
other		−.231*** (0.054)	.0013 (0.077)	−.0211 (0.076)	.0342 (0.066)
ability/family controls					
peabody intelligence			.0194*** (0.001)	.0179*** (0.001)	.0172*** (0.001)
HH income			6.1e−04** (0.000)	5.9e−04** (0.000)	9.0e−04*** (0.000)
single HH			−.156*** (0.020)	−.154*** (0.021)	−.122*** (0.021)
welfare			−.0781 (0.051)	−.065 (0.051)	−.0662 (0.046)
parents high school[(c)]			.0506 (0.048)	.0402 (0.048)	.0135 (0.039)
parents some college			.14*** (0.048)	.125*** (0.046)	.0927** (0.043)
parents college grad			.234*** (0.054)	.217*** (0.052)	.185*** (0.043)
parents > college			.361*** (0.059)	.35*** (0.056)	.322*** (0.049)
age fixed effects	*no*	*yes*	*yes*	*yes*	*yes*
grade fixed effects	*no*	*no*	*no*	*yes*	*yes*
grade*age fixed effects	*no*	*no*	*no*	*yes*	*yes*
school fixed effects	*no*	*no*	*no*	*no*	*yes*
observations	8385	8379	5774	5774	5774
r^2	.0056	.109	.246	.26	.313

Note: Standard errors in parentheses allow for clustering at the school level. Columns (1)-(5) are OLS regressions with dependent variable 'High School GPA'. Reference categories are: (a) health excellent, (b) white, (c) parents' education less than high school. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

One should be wary of causal interpretations that are based on such correlations between self-views and educational outcomes. Problems of unobserved heterogeneity and especially reverse causality remain. For example, self-esteem might affect success in school, but success in school is likely to increase self-esteem. This reverse causation will induce an upward bias in the self-esteem coefficient which cannot be directly addressed with a control variable strategy. At the same time, there are background factors that are likely to favor a positive self-image in students as well as to positively affect educational outcomes, e.g. parental care or ability. These factors are difficult to hold constant, even with a relatively broad range of controls like the ones employed in Table 1.

An often proposed strategy to infer causation, not only in self-esteem regressions but also in the broader literature on personality features in economics, is to use psychological variables measured prior to the predicted outcome. Several papers use child or early youth self-esteem (or personality measures in general), which is assumed to be independent from subsequent outcomes such as later schooling or labor market choices (Feinstein, 2000; Nyhus and Pons, 2005; Osborne Groves, 2005; Waddell, 2006; Mueller and Plug, 2006; Fortin, 2008; Heineck and Anger 2010; Drago, 2011). Under certain circumstances, the use of lagged measures may not do away with reverse causality: In our case, early self-esteem may be driven by individuals who anticipate to do well in school later in life, and similarly even pre-school self-esteem may already be boosted by good 'performance' in kindergarten, which in turn is possibly correlated with future school performance – a similar argument can be made for unobserved factors that drive early self-esteem and that may at the same time drive educational outcomes later in life. Such strategies that change the time frame from which psychometric measures are collected are important for research on personality. We contribute another step to this effort by proposing a different identification strategy which relies on exogenous variation in self-esteem caused by a quasi-natural experiment, i.e. the occurrence of adolescent acne.

3 An exogenous variation in self-esteem

The main contribution of this paper is to provide causal evidence on how self-esteem affects educational performance. To this end, in an instrumental variable approach we exploit exogenous variation in self-image generated by a quasi natural experiment that adolescents are faced with: the occurrence of skin problems due to acne. In the course of this chapter we support our claim that acne is a valid instrument. We start by giving evidence that – in line with the psychological literature – skin problems have an impact on self-evaluations in our sample. We then move on to argue that acne meets the exogeneity condition: First, we discuss the predispositional and behavioral determinants of acne as identified by the medical literature and conclude that having acne in puberty can be considered as good as randomly assigned, given certain covariates. Second, we also address the possibility that acne might affect grades through channels other than self-esteem, notably via discrimination by others.

3.1 Acne and its effect on self-esteem

The first condition that must be met for acne to be a valid instrument is that it needs to have an effect on self-esteem. Even though acne does not lead to any physical or cognitive impairment, the medical literature stresses that it is the psychosocial effects that make it more than a trivial disease. It mainly affects visible skin parts, and this variation in appearance may have especially severe psychological effects because "acne peaks in teenage years, a time crucial for building confidence and self-esteem" (Williams et al., 2012).[6] Not only do acne sufferers have to cope with a variation in looks that can affect self-esteem, but on top of this they can also be affected by the stigma that comes from the belief that acne is a result of lifestyle choices. In an anonymous leading article in the British Medical Journal, the author stated that acne is regarded as a "telltale mark", an "outward sign of moral

[6] While it is true that acne does not persist beyond adolescence in most individuals, Tan et al. (2001) find that only half of acne patients believe that there is a cure for the disease – which may exacerbate negative self-evaluations due to a feeling that one may have to suffer from the condition for a long time, possibly indefinitely.

defilement" (N.N., 1976). While this statement may sound antiquated, even nowadays it is widely held that individual behaviors such as poor hygiene, not getting enough sunlight exposure, or an unhealthy diet of e.g. chocolate and potato chips are responsible for the development of acne. Magin et al. (2005) provide a survey of the current medical literature that debunks such beliefs as myths and misconceptions, nonetheless they are common among the general population, acne sufferers themselves (Tan et al., 2001), and even among medical students (Green and Sinclair, 2001). In the face of these persisting myths, acne patients may feel that they themselves are responsible for developing acne – causing feelings of shame or even guilt that can aggravate any initial effects that the disease has on self-esteem.

Accordingly, Mallon et al. (1999) show that the self-esteem of acne patients is lower than that of control populations. The magnitude of impairment for these individuals is quite remarkable: patients with severe acne perceive their quality of life to be as low as what is usually reported by chronic asthma, diabetes and epilepsy sufferers. Similarly, and closer to the outcome we use, Dalgard et al. (2008) report more negative self-evaluations in acne patients, and Thomas (2005) also attributes negative self-images to acne.

In order to assess whether such a pattern of reduced self-esteem in those with acne is present in our data as well, we use the AddHealth question on skin problems. It reads: "In the past 12 months, how often have you had skin problems, such as itching or pimples?". Note that the wording does not explicitly mention acne (even though the question is labeled "acne" in the AddHealth data). This is not a unique feature of the skin problems question. Quite the contrary: whenever AddHealth inquires about health in the "General Health" module the questions refer to symptoms rather than specific conditions. Presumably, this is because adolescents would have had to be diagnosed with a disease in order to state that they are suffering from it, whereas symptoms can be described even in the absence of knowledge about any underlying conditions. Regardless of why the question does not mention acne, there are several reasons which lead us to believe that it does not measure anything other than acne. For one, acne is the most important reason why individuals develop pimples. Even in adults pimples are almost exclusively associated with acne, but remember that the

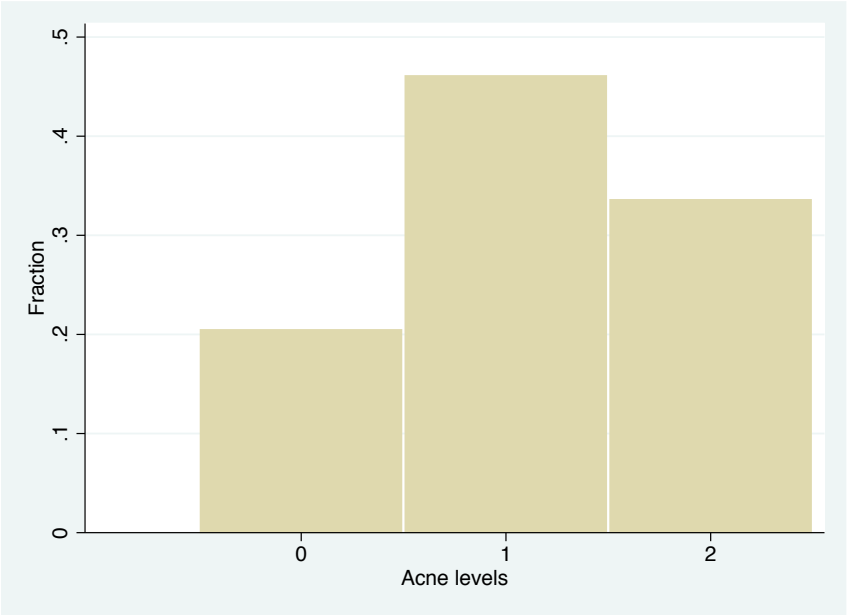
individuals in our sample are at an age where it is even more reasonable to assume that those who state to have pimples will actually be affected by acne.[7] This idea is supported by the fact that the stated skin problems in our sample are correlated with age, a finding that is highly indicative of acne and would not be expected with other conditions. In addition, we checked the prevalence of "skin problems" in our sample against the prevalence of acne that the medical literature reports. In the original AddHealth variable 15% report having skin problems "almost every day" – fitting the prevalence of moderate to severe acne in adolescents, which Williams et al. (2012) report to be 15-20% in individuals between age 15 and 17. Similarly, the numbers frequently cited in the medical literature state that around 80% of adolescents suffer from some degree of acne (Halvorsen et al., 2008; Tan et al., 2007; James, 2005) which is very much in line with our data, where roughly four out of five respondents state to have skin problems of some degree. Against this backdrop, even though acne is not mentioned, asking adolescents about pimples clearly alludes to this condition and with respect to the part of the question which inquires about "pimples" we have no problem interpreting it as being about acne.

On the other hand "itching" is also explicitly mentioned in the AddHealth question. Despite the fact that itching probably does not come to mind as an acne symptom, medical research has found that it is actually quite common in acne patients (Reich et al., 2008; Lim et al., 2008), and therefore the AddHealth question can be thought of as mentioning the two most important acne symptoms. Yet itching in some respondents could also be caused by conditions other than acne, the most common of which are probably allergies. With regard to this objection, first note that unless allergies and acne are highly correlated, i.e. unless those with allergies are actually a subset of those with acne, we would not expect to find the matching acne prevalence figures we just described. Rather the share of those experiencing skin problems from allergies *or* acne at least a few times should be higher than the 80% we observe. Furthermore, if the skin problems question also captures that some people have allergies, controlling for allergy should actually improve the strength

[7]Other causes of acne resembling symptoms are quite rare. One alternative driver of acne resembling symptoms such as pimples is dioxin exposure (Passarini et al., 2010). Because this mainly occurs in industrial workers, it is not of concern for our sample.

of the instrument in the first stage, as it seems likely that allergies do not have much of an effect on self-esteem. For some of the estimations we actually have a control variable available that explicitly inquired with the parents about any allergies that the adolescent might have.[8] In fact, where we have this information (only in Wave I), self-esteem levels of individuals with allergies are not statistically different from those respondents who do not have allergy, and adding allergy as an additional control variable leaves our IV results unchanged. Taken all this evidence together, we are confident that we can effectively rule out that answers to the AddHealth question are to a notable extent driven by other skin conditions – so we interpret the question as being about acne and from here on we will use the terms "skin problems" and "acne" interchangeably.

Figure 3: Distribution of acne levels, Wave I.



Note: The figure compares the observed distribution of acne levels (0=never, 1=less than weekly, 2=at least weekly). The sample is the same as the estimation sample used in Table 2. Number of observations $N = 4927$.

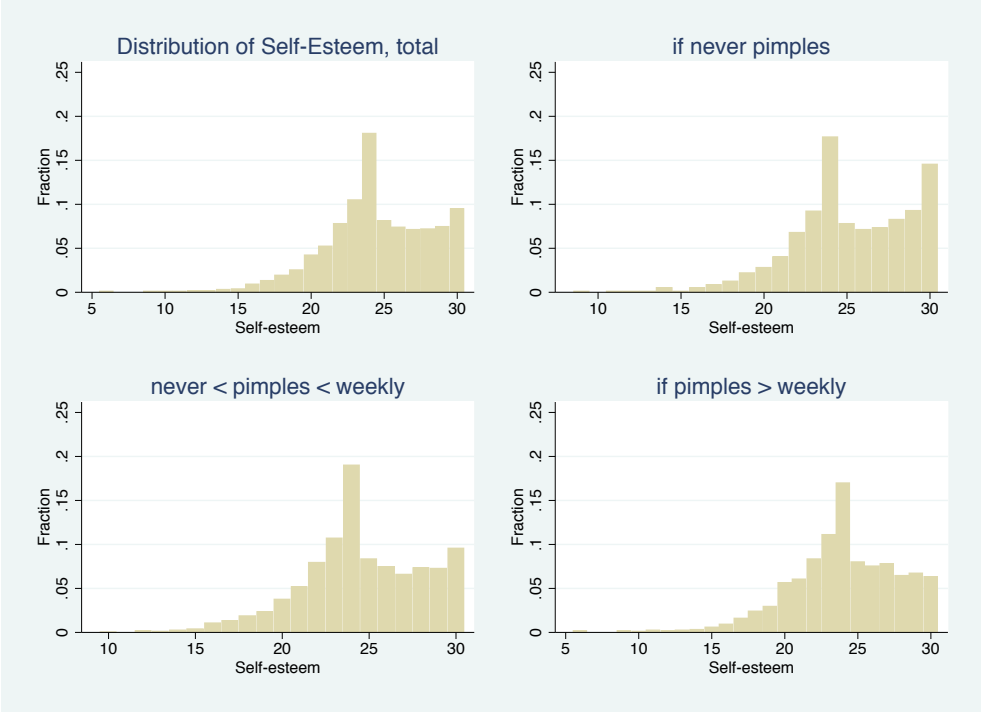
Figure 3 shows the distribution of the acne variable values in our sample (Wave I). This variable is a modification of the original AddHealth question – using the original five answer categories "never", "just a few times", "about once a week", "almost every day",

[8] In addition to the student questionnaire, there is also a parental questionnaire which actually did inquire about some specific diseases. Unfortunately it does not include a question on acne.

and "every day" lumps about 70% of respondents in just two answer categories, and so we sum up the highest three categories to make the distribution less skewed. The answer categories of our question about the frequency of skin problems then are "never", "a few times", "at least weekly". As already noted, most students had skin problems a few times in the past year, followed by those who have problems at least on a weekly basis, and the smallest group is the one that states never to have had skin problems at about roughly 20%.

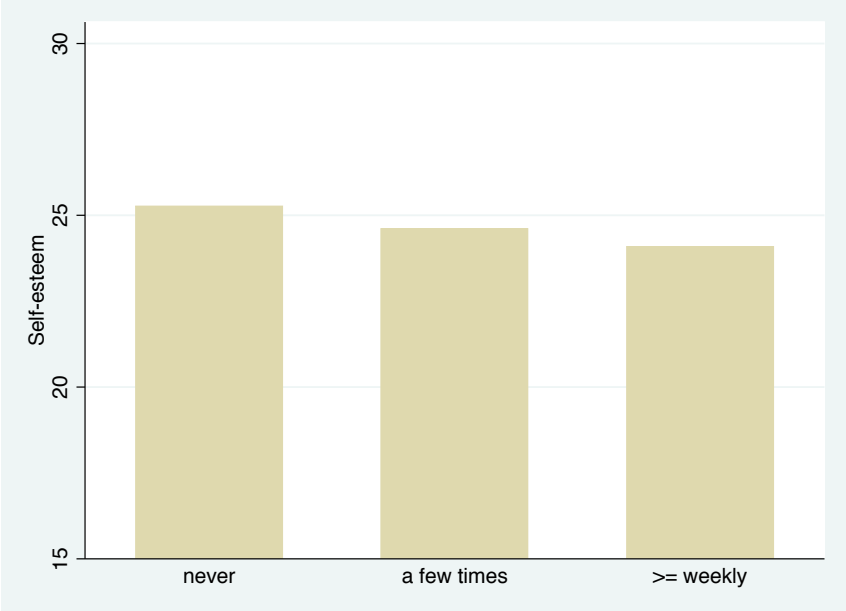
Provided that this skin problem variable is a good measure of acne, we can now evaluate whether it meets the first condition for being a valid instrument, i.e. that skin problems are associated with lower self-esteem levels. Corresponding to the evidence that the medical literature has gathered, a visual inspection of our data shows that skin problems are linked to lower self-esteem levels in the AddHealth data. Figure 4 illustrates that the distribution of self-esteem shifts to the left (i.e. towards lower self-evaluations) with increasing levels of skin problems, and plotting self-esteem against acne levels in Figure 5 also supports our claim that higher levels of acne are associated with lower average levels of self-esteem. To put the magnitude of self-esteem differences across the categories of skin problems into perspective, the difference in mean self-esteem between those in the 'never' and those in the 'at least weekly" acne category is slightly more than one Rosenberg point – this one point represents roughly one third of a standard deviation in self-esteem. It is also roughly the difference in means of self-esteem between AddHealth respondents who assess their general health as "good" and those who rate it as "fair", i.e. going from category three to category two on a five point health scale. These descriptive statistics already support our claim that acne has a fairly important effect on self-esteem, yet ultimately what matters is that the instrument has a strong enough effect in the first stage of our estimations. In chapter 4.1 our estimates will show just that: we are far from having a weak instrument problem.

Figure 4: Distribution of self-esteem, Wave I.



Note: The figure compares the observed distribution of self-esteem by levels of acne. The top left graph displays the overall distribution. The sample is the same as the estimation sample used in Table 2. Number of observations $N = 4927$.

Figure 5: Self-esteem and acne, Wave I.



Note: The figure displays the mean self-esteem levels by levels of acne. The sample is the same as the estimation sample used in Table 2. Number of observations $N = 4927$.

3.2 The pathogenesis of acne and exogeneity

Making a case that acne meets the second instrument condition is of course much harder: the exogeneity requirement demands that acne is as good as randomly assigned and that

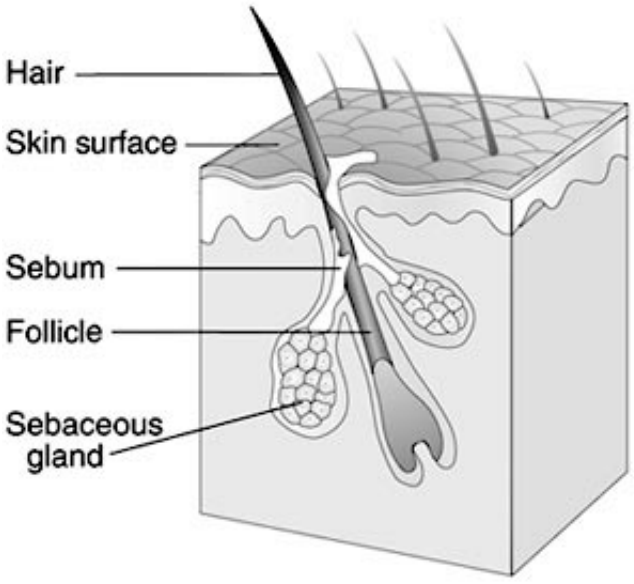
it can only be linked to grades through self-esteem, conditional on covariates.

The medical literature lends credibility to the idea that the onset of acne in puberty can be considered as good as randomly assigned with respect to academic outcomes. Four factors have been identified to be the main causes of acne.[9] First, sebum (oil) production matters in the pathogenesis of acne. A pimple is basically just an oil filled skin lesion, and this means that the sebaceous glands in the skin need to produce oil, otherwise an acne lesion cannot develop (Figure 6 shows a hair follicle with attached sebaceous glands). Sebum production is closely tied to androgen production, which is the second factor. Androgens play a role because the onset of androgen production in puberty leads to a growth in the sebaceous glands which in turn produce oil. Because generally the oil can just travel from the sebaceous gland through the hair follicle to the skin surface without causing acne lesions, a third factor is needed in order for sebum production to become a problem. This is where hyperkeratinization comes in: the inner lining of the follicular duct is renewed in certain intervals, and the dead skin cells are then transported to the skin surface. This is a normal process, yet hyperkeratinization leads to it being accelerated and thus generates large amounts of dead skin cells, which create potential for clogging the follicular duct. When this happens, oil cannot leave the follicle and an acne lesion develops (it is easy to see in Figure 6 how a clogging of the follicle could create an acne lesion or pimple). The fourth factor is propionibacterium acnes, a bacterium which resides on the skin surface of every human and thrives in clogged follicular ducts thus aggravating acne.

This excursion into the pathogenesis of acne serves to show that acne is caused by local processes in the skin (hyperkeratinization, sebum production) and these are per se not under suspicion of affecting any hard skills which may drive grades – such as memory, logical reasoning, i.e. cognition. In the same vein, acne does not affect physical ability which may contribute to school grades. Put differently, acne sufferers are still able to walk straight and think straight, and in this respect, acne should be statistically independent of school outcomes. Similarly, it is important that in twin studies the processes responsible

[9]See Zouboulis et al. (2005), Williams et al. (2011), or Kurokawa et al. (2009) for a summary of the pathogenesis of acne.

Figure 6: Hair follicle.



Source: Wikimedia Commons. `http://en.wikipedia.org/wiki/File:Hair_follicle-en.svg`.

for acne have been shown to be mostly driven by genetics (Bataille et al., 2002). Because there is no evidence linking these genetic drivers of local skin processes to factors such as genetic variation in ability, we can assume the onset of acne to be orthogonal to this unobserved inherent driver of educational performance. In other words, neither does acne affect ability in those who have the disease, nor are people of a certain ability more likely to develop acne.

A few important caveats and clarifications are in order regarding this general statement is that, as just noted, only after puberty sets in, the human body starts producing the hormones which are a prerequisite to developing acne, and so we will observe a correlation between developmental or pubertal status and acne levels. Using acne as an instrument without factoring out pubertal status could then violate the exogeneity condition to the extent that those who are more developed at a certain age are more likely to have acne, and the fact that they are more mature than their peers of the same age may at the same time facilitate their obtaining good grades in school. To guard against this possibly confounding factor, we additionally net out maturity levels by adding a variable that asked interviewers how physically mature the respondent was compared to other respondents his age, and so we essentially compare individuals that are at the same maturity level and therefore

pubertal stage.

Because androgen production plays a role for acne, one could also suspect that androgen *levels* might be a problem for our identification strategy: androgens might facilitate or hinder the obtaining of good grades, e.g. by influencing aggressiveness. Here it is important to realize that only after a certain threshold level of androgens is reached in puberty the body starts producing significant amounts of oil. When this has happened acne can occur, yet circulating androgen level above this threshold do not correlate with acne severity – in fact, most acne patients have normal levels of circulating androgens.[10] We already include controls for pubertal stages in our estimations and this factors out whether an individual has already reached the threshold hormone level. Any remaining heterogeneity in androgen levels that might affect grades should by this argument not be related to whether an individual has acne. Furthermore, in chapter 4 we conduct extensive falsification tests which substantiate this notion that acne does not measure diversity in either circulating hormones or physical development.

Regarding the final acne factor, propionibacterium acnes, it should be stated that the existence of this bacterium on the skin is not linked to poor hygienic habits which some might construe to be a function of socioeconomic status. Rather the bacterium resides in the skin of every human being and can therefore be assumed to be independent of educational performance. In fact, there is no evidence that links the occurrence of acne to social status or status related behaviors. Nonetheless, our specifications include a battery of controls that capture heterogeneity in this respect: household income, parental education, whether the household is on welfare, and whether the respondent lives with a single parent.

Since the occurrence of acne is mostly driven by genetic factors, individual behavior has only a marginal effect on whether a person develops acne. Specifically, as noted above, acne in general is not a disease of poor hygiene or lack of cleanliness (Webster, 2001). Furthermore, there is little evidence to support the idea that dietary habits cause acne

[10]This is why before having reached puberty, acne does not occur and therefore androgens are said to play a "permissive role" in the pathogenesis of acne (to be more precise, before adrenarche, acne does not occur. The onset of adrenarche marks the start of androgen production and can preceed what is known as gonadal puberty). The increased sebum production in acne sufferers is thus thought to be due to local enzymes in the skin or an extreme responsiveness of the sebaceous glands to androgens (Kurokawa et al., 2009).

(Davidovici and Wolf, 2010; Williams, 2012). This is why a change of diet is typically not mentioned as a treatment option, and the same is true for a change of hygiene habits (James, 2005; Williams et al., 2012). One of the oldest acne myths probably is that chocolate consumption plays a role in causing acne. Fulton et al. (1969) conducted a study where the treatment group had to eat chocolate bars, and found that it did not contribute to acne. Despite the "myth" status of diet as a cause of acne, we will add controls for body mass index (BMI) and diet. In Wave I of our data we measure an unhealthy diet by a question on whether the respondent usually eats snack foods for breakfast, and in Wave II respondents declared whether they had eaten chocolate or chips on the day before the interview (the snack foods question was not asked in Wave II).

The use of oral contraceptives on the other hand suppresses androgens and sometimes the birth control pill is even prescribed as an acne treatment. We will thus factor out whether the respondent is on the pill.[11] To further account for the availability of treatment options, we also control for whether in the past year there was at least one instance where the respondent was unable to see a doctor even though they wanted to do so. Selection into medical treatment might also be driven by an "indifference to looks" which in turn may be positively correlated with grades, such that kids from a family with a preference for beauty may be less supported by their parents when it comes to academics and consequently perform worse. AddHealth does not provide a direct measure of such attitudes, yet we have some information on parental priorities: parents were asked whether their top priority for their offspring is that they be excellent students rather than being popular or a great athlete. In order to guard against such priorities driving our result, we will add this variable as a control.

This excursion into the pathogenesis of acne has shown that the event of developing such skin problems should be independent of potential educational outcomes, after adjusting for the covariates mentioned above – in that sense acne can be considered as good as randomly assigned. Yet for exogeneity to hold we need to make sure that acne does not affect school performance through a channel different from self-esteem. Specifically,

[11] This variable is coded zero for male respondents.

because acne constitutes a variation in looks, it may lead to others treating those with acne differently, i.e. acne may induce outright discrimination. Other adolescents may pick on those with skin problems, and as a result acne sufferers may perform worse in school. Much of this "discrimination by other adolescents" effect probably runs via reduced self-esteem: kids pick on those with acne and this reduces the self-esteem of acne sufferers, which may then affect performance of students in school. In this case the exclusion restriction holds without further controls. On the other hand discrimination need not run exclusively via acne reducing self-esteem: school performance could be affected if those with acne were actively being distracted by others when in class, or if they were kept out of study groups because they have acne. It is also possible that teachers discriminate against students with acne, and grade them more harshly – thus violating the exclusion restriction. To address these concerns we add controls that account for possible discrimination: how often the respondent had trouble with teachers, how often the respondent had trouble with students, and whether the respondent feels that students at their school are being treated fairly by teachers. Using this setup, in the next chapter we present our estimates of the effect of self-esteem on performance in school and compare it to the OLS results.

4 Results

In this section we provide estimates of the causal effect of self-esteem on task performance obtained from instrumental variable estimations. We start our analysis with a cross-sectional model which explains between-student variation in school grades. Given the identifying assumptions of the previous chapter these baseline estimates pinpoint the causal path from self-evaluations to school performance. However, to empirically substantiate our claim that acne is an appealing choice as an instrument, we will also provide robustness checks which take advantage of the panel nature of our data, and furthermore we will conduct a number of falsification tests.

4.1 The effect of self-esteem on school performance

Table 2 lays out the main IV-results. Our specification expands the most comprehensive OLS specification by adding the controls needed for the identification strategy. These comprise indicators capturing how physically mature the respondent is compared to other respondents their age, and we condition on whether the respondent felt that teachers at their school treat students unfairly, as well as whether the respondents reports having trouble with teachers or other students in order to net out discrimination effects. As controls for diet and eating habits, we include an indicator for whether the respondent regularly eats snack foods for breakfast as well as self-reported BMI. To take into account availability of medical treatment, we control for whether the respondent was unable to see a doctor in the past year even though he wanted to. To the extent that the socio-demographic controls (parental education, household income etc.) do not already capture the background factors which shape the respondent's tendency to accumulate human capital, we use information on whether parents stated that their highest priority is for their offspring to be a brilliant student.

In the previous chapter we have already shown that there is a strong correlation between acne severity and self-esteem levels, suggesting that acne meets the first condition for being a valid instrument. In Table 2 the first stage estimates confirm this: they suggest that skin problems are a good predictor of self-esteem, even after conditioning on covariates (the full first stage is shown in Table 10 in the Appendix). The coefficient on acne is negative as expected and therefore supports the stylized fact that acne reduces self-esteem. Furthermore, the acne coefficient is statistically highly significant and the instrument F-value of about 26.5 clearly exceeds the critical value for weak instruments in the just identified case.

In addition to having an effect on self-esteem, acne should also display a significant coefficient in the reduced form regression of high school grades on acne. Interestingly, a comparison of means across acne levels (see Figure 7 in the Appendix) shows that those with the most severe skin problems are the ones that on average attain the best grades, and

Table 2: IV estimates, Wave I.

	IV 2nd stage	
self-esteem	−.202***	(0.062)
younger than most$^{(a)}$	−.12*	(0.066)
younger than some	−.0747	(0.080)
older than some	−.0956**	(0.040)
older than most	−.0132	(0.052)
no medical treatment	−.122**	(0.057)
pill	−.0104	(0.094)
BMI	−.0093**	(0.004)
snack food	−.0658	(0.060)
important brilliant student	−.0065	(0.033)
trouble w/ teacher	−.201***	(0.030)
trouble w/ students	−.105***	(0.028)
teachers unfair	−.158***	(0.031)
all OLS controls	*yes*	
age fixed effects	*yes*	
grade fixed effects	*yes*	
grade*age fixed effects	*yes*	
school fixed effects	*yes*	
	1st stage	
acne	−.298***	(0.058)
1st stage F-val	26.5	
	Reduced form	
acne	.0602***	(0.015)
observations	4927	

Note: Standard errors in parentheses allow for clustering at the school level. IV estimates for Wave I. The endogenous variable 'self-esteem' is instrumented with 'acne level', the dependent variable is 'High School GPA'. The middle and bottom panel show 1st stage and reduced form coefficients for the instrument 'acne level'. Reference categories are: (a) looks neither younger nor older. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

the reduced form estimation presented in Table 2 paints the same picture: A positive and highly significant reduced form coefficient establishes that those with worse acne actually perform better in school (for the full reduced form see Table 11 in the Appendix). Because those with the heaviest skin problems are also the individuals with the lowest self-esteem, these results already give some indication that having high self-esteem may actually have negative effects on performance in school – contrary to what is typically found in the literature, and going against the popular belief that higher self-esteem always brings about better outcomes.

The second stage results in Table 2 support this hypothesis: when using the variation in skin problems as an instrumental variable the coefficient of self-esteem switches signs in comparison to OLS, revealing a statistically highly significant negative causal effect of self-esteem on high school grade point average. The magnitude of the coefficient is non trivial and suggests that a one standard deviation increase in self-esteem causes a .6

standard deviation decrease in GPA. On a general level this is in line with the view that non-cognitive factors can have huge effects on economic outcomes. However, our result suggests that in the case of self-evaluations, the effects are quite different from of what is usually assumed: when endogeneity of self-views is taken into account, self-esteem is negatively related to school performance. This not only confirms our suspicion that the OLS coefficients are biased upwards, but it suggests that this bias is severe enough to even produce the "wrong" coefficient sign.

4.2 Robustness checks

Before we get to discussing mechanisms that may generate a negative effect of self-esteem on school performance, this and the next section are dedicated to robustness checks and falsification tests. The results presented so far were obtained using data from Wave I of AddHealth only. We will now present some robustness checks in order to further endorse our claim that the cross sectional setup we have used in the previous chapter already provides a reliable estimate of the causal effect. We show that using longitudinal data and dropping the assumptions specific to cross-sectional estimation gives qualitatively similar results.

As a criticism of our identification strategy, it might be put forward that the control variable set does not sufficiently take into account heterogeneity in personal traits or features. For example, the parental priority measure does not directly control for 'indifference to looks' or 'vanity' - individual characteristics that may be related to the willingness to accept visible skin shocks as well as to intrinsic interest in academics. Likewise, highest degree of parental education and the Peabody Picture Vocabulary Test might be imperfect controls for cognitive ability. Although there is no medical evidence whatsoever that acne is related to ability, one may raise the objection that both acne and intelligence have been shown to be greatly affected by genetic factors, and thus there is no way to completely rule out that the genes responsible for being affected by adolescent acne covary with 'ability genes' that lead to better grades. A related point is that dynastic effects in self-esteem may

play a role: it is conceivable that parents with a higher genetic predisposition towards skin problems have been exposed to an acne-induced self-esteem shock in their youth themselves. Assuming that these shocks to self-esteem persist, the parents then pass on to their children both a genetic predisposition to acne, and a low self-esteem environment which may have a direct effect on grades. In this case, acne will be correlated with the unobserved 'low self-esteem background', and the IV self-esteem coefficient will then pick up direct effects of this environmental self-esteem factor, too.[12]

To assess whether these fixed characteristics bias our cross-section IV results, we exploit the panel structure of the AddHealth survey and run first differenced IVs between Wave I and II. First differencing of course removes all common variation of acne with individual characteristics that are time-fixed between the two waves – among them the above mentioned individual abilities, individual tastes for looks, parental background and dynastic effects. If this specification delivers results similar to what we have obtained from the cross-sectional approach, we can be confident that the results from our main specification in chapter 4.1 do not suffer from problems of the instrument being correlated with unobserved time fixed heterogeneity.

Column (1) of Table 3 show the results using first differenced data. We regress differenced school grades on differenced Rosenberg self-esteem, which is instrumented by differenced levels of acne. As controls, we use all cross-section controls that are available for both waves. Because some of the variables are not available in Wave II and because the time invariant variables drop out, we here have a reduced number of controls. The coefficient of interest in the first differenced IV approach is -0.169, i.e. similar to that of the cross-section IV coefficient of -0.202. It should be noted that the significance is somewhat poorer since the within variation in acne over such a short time horizon of roughly one year is smaller than the variation between individuals. In fact, for most interviewees the time between the first and second wave interview was less than one year, making it even harder to obtain significant results. Accordingly, the F-value on the excluded instrument is rather low and does not reach the often used rule of thumb of a critical value of 10.

[12]To the extent that parents with skin problems have a higher probability to mate this will be more likely.

Table 3: First differenced IV and OLS estimates.

	(1)		(2)	
	FD-IV 2nd stage		***FD-OLS***	
self-esteem	−.169*	(0.101)	.0107***	(0.003)
health	−.0876**	(0.038)	−.0161	(0.010)
single HH	−.117**	(0.046)	−.0678***	(0.025)
welfare	.0434	(0.061)	−.0382	(0.032)
bmi	−.0055	(0.008)		
no medical treatment	−.0343	(0.038)		
phys maturity	−.0127	(0.013)		
trouble teacher	−.0928***	(0.021)		
trouble students	−.0433**	(0.017)		
unfair teacher	−.0547**	(0.022)		
observations	4916		5324	
	FD-IV 1st stage			
acne	−.141**	(0.064)		
instrument F-val	4.79			
observations	4916			
	FD-IV Reduced form			
acne	.0238**	(0.011)		
observations	4916			

Note: Standard errors in parentheses allow for clustering at the school level. Column (1) is a first differenced IV estimation, where the endogenous variable 'self-esteem' is instrumented with 'acne level'. Column (2) is a first differenced OLS regression. The dependent variable is 'High School GPA' in both specifications. For the IV specification, the middle and bottom panel show 1st stage and reduced form coefficients for the instrument 'acne level'. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

However, since we are using a single instrument for a single endogenous variable and just identified IV is approximately median unbiased, this is much less of a problem than in an overidentified setup (Angrist and Krueger, 1999; Angrist and Pischke, 2009). In addition, Anderson and Rubin (1949) tests can be used for inference with weak instruments (Dufour and Taamouti, 2005; Andrews and Stock, 2005). In our case, by using this test which is robust to the presence of weak instruments, we can at the 5%-level reject the null that the coefficient on self-esteem is zero. This reconfirms the existence of a negative effect of self-esteem on school performance, even though the self-esteem coefficient in the first differenced IV framework is only significant at the 10%-level. The mere fact that the point estimates in the first differenced framework are similar to the cross section adds credibility to the claim that time-fixed unobservables are more potential than actual problems and do not much contaminate our IV results presented in section 4.1.

The relative invariance of our IV self esteem coefficient to differencing out fixed factors contrasts with how differencing alters results in the standard OLS framework. In Table 3 we also report the non-instrumented first differenced results. As can be seen from Column

(2), non-instrumented self-esteem is still positively related to school grades, but compared to the cross sectional OLS estimates the coefficient is cut down by about 40 percent (the coefficient drops from 0.0177 to 0.0107). This empirically backs our suspicion that simple OLS estimates are biased upwards due to time fixed unobserved heterogeneity.

First differences remove time-fixed unobservables, but the instrument may still be invalid due to backwards causality: It might be that acne is not causally related to school performance, but that school performance is conducive to acne. In what follows we argue that this is not likely to drive our findings, either.

In particular, one may believe that doing badly in school causes stress which in turn can then cause acne. Note first that the medical evidence does not suggest that stress causes acne and therefore does not support such a story to begin with. Second, in the reduced form regressions of GPA on acne we obtain a positive acne coefficient – this is not in line with bad grades (negative stress in school) producing skin problems, because in this case one would expect to find a negative relationship between acne and GPA. Nevertheless, we can construct a story where good rather than bad grades may actually cause stress: Doing well in school may lead to bullying by others, and one may believe that such social distress may cause acne. This scenario would indeed produce a positive acne coefficient in the reduced form regressions. We do, however, already condition on whether students have trouble with their social environment which should hold this channel constant; and still, we come up with a positive coefficient in the reduced forms.

In addition to these arguments, we also employ a specification that includes the lagged dependent variable. This addresses issues of reverse causality not only due to stress but also on a more general level: Assume that for *whatever* unknown reason school performance produces acne. If this is why our instrumental variable generates the coefficients shown previously, then our results should disappear when conditioning on past grades. But this is not the case: In column (1) of Table 4, we report the IV-results from Wave II controlling for Wave I school grades. Lagged GPA is a very strong predictor of current GPA suggesting that there is inertia in school success, yet we obtain similar results to the cross-section

Table 4: IV and OLS Estimates: Lagged Dependent Variable Models (Wave II).

	(1)		(2)	(3)
	LDV-IV 2nd stage		***LDV-OLS***	***OLS***
self-esteem	−.153**	(0.071)	.0114***(0.003)	.0238***(0.004)
GPA lagged	.73***	(0.028)	.715*** (0.022)	
younger than most	.0029	(0.052)		
younger than some	−.0366	(0.035)		
older than some	−.003	(0.036)		
older than most	.132*	(0.070)		
no medical treatment	−.0746	(0.052)		
pill	−.007	(0.049)		
BMI	.001	(0.003)		
chips	.0376	(0.029)		
trouble w/ teacher	−.118***	(0.020)		
trouble w/ students	−.0673*	(0.036)		
teachers unfair	−.0697***	(0.027)		
all OLS controls	*yes*		*yes*	*yes*
age fixed effects	*yes*		*yes*	*yes*
grade fixed effects	*yes*		*yes*	*yes*
grade*age fixed effects	*yes*		*yes*	*yes*
school fixed effects	*yes*		*yes*	*yes*
	1st stage			
acne	−.233***	(0.066)		
1st stage F-val	12.4			
	Reduced form			
acne	.0356***	(0.012)		
observations	3684		3760	3813

Note: Standard errors in parentheses allow for clustering at the school level. Column (1) is an IV estimation for Wave II, where the endogenous variable 'self-esteem' is instrumented with 'acne level'. Column (2) and (3) are OLS regressions for Wave II. The dependent variable is 'High School GPA' in all specifications. The middle and bottom panel show 1st stage and reduced form coefficients for the instrument 'acne level'. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

IV: the coefficient on instrumented self-esteem is significant and at −0.15 still of sizable magnitude.[13] For reference, column (2) of Table 4 shows the non-instrumented version of the lagged dependent variable model, and column (3) reports the OLS results from the baseline specification for Wave II. Both these models yield a positive self-esteem coefficient, but the inclusion of the lagged dependent variable reduces it by more than one half.

The sum of all these robustness checks provides evidence that the self esteem point estimate obtained from the cross-sectional IV is not driven by either unobserved time-fixed factors or by reverse causality.

[13]The set of controls is somewhat different from the Wave I cross sectional specification: The Wave II questionnaire does not inquire about whether the respondent eats snack foods for breakfast. It does however include the question "Yesterday did you eat potato chips, corn chips, tortilla chips, pretzels, or popcorn?", and we use this variable in order to capture unhealthy diet. In addition, there was no question on household income in Wave II and the verbal intelligence test was only administered in Wave I. In both these cases we use the Wave I value.

4.3 Falsification tests

As we have said before, the occurrence of acne is closely tied to the onset of puberty in the sense that pimples do not develop before puberty. Because puberty may also affect school performance, in our instrumental variable estimations we control for whether the respondent has already reached puberty by including information on their age and physical development. Now, if advanced pubertal staging has a direct positive effect on school performance (not through acne) and we are not able to fully capture the pubertal stages with the above mentioned controls, then our results may be due to the acne variable picking up these effects. In a similar fashion one may believe that androgen levels directly influence school performance, and even among those individuals who have already reached puberty, the acne variable may simply measure heterogeneity in circulating androgen (despite the fact that the medical literature tells a different story; see section 3). First, it is worth observing that testosterone is believed to be associated with adverse behaviors like non-cooperative or aggressive tendencies and it would seem that these would actually go with worse school outcomes – which does not fit the positive coefficient we observe in the reduced form estimations. Still, one may conversely hold the belief that it is exactly this aggressiveness that somehow enables those with high testosterone levels to actually obtain better grades, and this is what drives our results.

In what follows, we provide a number of falsification tests. In these tests we use alternative instruments that are known to be highly correlated with both pubertal staging and androgen levels. If indeed the above mechanisms spuriously generate the negative self-esteem coefficient, then the alternative instruments should deliver similar results. We show that this is not the case.

Direct information on testosterone levels or other androgen activity is not available in AddHealth. But in a number of falsification tests we can use variables other than acne, which are typically used to determine the pubertal stage of an individual and are affected by androgen levels. For men we use as placebo instruments the amount of hair the respondent has under his arms, the extent to which the respondent's voice is lower than in grade school,

Table 5: Falsification tests.

	(1) GPA		(2) GPA		(3) GPA		(4) GPA	
Falsification Wave 1, 2nd stage								
self-esteem	−.864	(3.681)	.105	(0.186)	.0303	(0.250)	−.0588	(0.132)
Falsification Wave 1, reduced form								
R has hair under arms	−.0162	(0.016)						
R has lower voice			.0106	(0.020)				
R has facial hair					.0031	(0.026)		
R has ever menstruated							−.0425	(0.097)
Original estimation Wave 1, 2nd stage on falsification sample								
Rosenberg	−.267*	(0.149)	−.273*	(0.154)	−.279*	(0.160)	−.161**	(0.072)
Original estimation Wave 1, reduced form on falsification sample								
acne	.0615***	(0.023)	.0614***	(0.023)	.0613***	(0.023)	.0593***	(0.022)
observations	2406		2411		2407		2510	

Note: Standard errors in parentheses allow for clustering at the school level. The placebo instruments for self-esteem in Columns (1)-(4) are puberty indicators of which it is known that they are correlated with hormone levels. Among these, 'R has hair under arms', 'R has lower voice' and 'R has facial hair' are available for male respondents only, whereas 'R has ever menstruated' is available for females. In the top and middle panel, results from the reduced form and second stage are shown. The bottom part of table also reports the coefficients on 'self-esteem' and 'acne levels' when our original IV specification (1) from Table 2 is estimated on these smaller (male and female only) samples. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

and the thickness of facial hair. For women we use a variable that measures whether the respondent has ever menstruated.[14] These alternative instruments are associated with testosterone and physical development, so if it is through this channel that acne spuriously produces our results, then we should see similar results when using these other variables as instruments for self-esteem.

We check this in Table 5. In line with our argument that acne does not capture androgen levels, columns (1)-(4) show that none of the alternative measures of physical development or maturity has an effect on GPA in the reduced form estimations. Similarly, the 2nd stage self-esteem coefficients very much depend on the instrument used and none of them is statistically significant. The samples for these estimations are much smaller than our main sample, as the falsification instruments were only asked of either male or female respondents and therefore cut sample size in half. In order to make sure that the insignificant results are not an artefact of the reduced sample size, we re-estimate the IV specifications with our original acne instruments on these smaller samples and check that this leads to results similar to our main specification. We still find a statistically significant negative effect of self-esteem on grades, so we can be confident that this is not the case. Overall, the falsification tests therefore strongly support our view that heterogeneity in developmental status or circulating hormone levels is not what generates our results.

5 Channels – does self-esteem alter behavior?

So far we have presented evidence that the negative relationship between self-esteem and grades holds across a number of specifications. Why higher adolescent self-esteem could bring about lower grades in school is a question we have not yet addressed. The standard reasoning in the literature is that positive self-evaluations makes individuals more confident in succeeding at a task and this makes them 'try harder', i.e. put in more effort which in turn improves performance. Following this idea self-esteem and effort are complements,

[14]The variable we have for women provides for a possibly less convincing falsification test than the variables we have for men, because in our sample of female high school students 95% had already reached menarche.

Table 6: Channels.

	(1) drive mls	(2) work hrs	(3) intercourse	(4) marijuana
Channels Wave 1, 2nd stage				
self-esteem	.22*** (0.074)	2.36*** (0.882)	.117*** (0.041)	.047 (0.039)
observations	3343	4909	4917	4901

Note: Standard errors in parentheses allow for clustering at the school level. All specifications are IV estimations as in column (2) of table 2. 'self-esteem' is instrumented with 'acne levels' in all specifications. The dependent variable is indicated in the respective column headers. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

but our finding that those with higher self-esteem perform worse does not fit this story. Quite the contrary, our results are consistent with the idea that self-esteem and effort can also be substitutes – the question is: can we find evidence for this? Unfortunately, our data does not provide detailed information on a respondent's daily time use. In particular, we do not know which share of their time endowment students devote to homework and learning. However, the Add Health questionnaire provides some information about leisure activities that may be considered as inversely related to time spent on academics and therefore inversely related to effort expended in school.

Table 6 reports the results from regressing some of these activities on instrumented self-esteem using Wave I data (controls are the same as in the baseline IV). In column (1) we see that among those who are of legal driving age the higher self-esteem individuals drive more miles per week in a car – which in the US is a pretty good proxy for having a social life, as in most places in order to get anywhere a car is needed.[15] Column (2) shows that higher self-esteem individuals also work more hours in a job after school during a typical school week. Those with higher self-esteem are also more likely to have had intercourse, as column (3) shows. When it comes to ever having used marijuana, we find no difference between those with high and low self-esteem (see column 4). We do not at all intend to make any judgements about whether any of these behaviors are good or bad uses of time, but rather see them as indicators for whether an individual is busy doing things outside the realm of academics. Overall, there is some evidence that this is the case to a larger extent for those with higher self-esteem.

These results indicate that changes in self-esteem alter the behavior of the students in

[15]The sample for this estimation includes only individuals who are at least 16 years of age – the legal driving age in the US.

our sample. This is in line with several explanations. First, it is likely that individuals with high global self-esteem – which the Rosenberg scale captures – also have confidence in specific capabilities, e.g., in their cognitive skills. The standard literature predicts that these people will try harder at a task they feel able to manage well, and we should thus expect them to have better grades in school. However, under reference point preferences things can be different: suppose that an individual has set a fixed goal to be attained at a specific task, with little utility gains from a performance that exceeds this reference point. Believing that this specific goal can be easily reached may then induce the individual to expend little effort at that task. To be specific, in our setup it is conceivable that one of the goals in adolescence is not to maximize human capital, but to meet a minimum goal, e.g., to graduate. Then, it is possible that – because they believe they will pass their exams anyway – individuals with higher confidence in their abilities exert less effort in school and turn to other tasks. As a consequence we should observe lower grades for these students.

A second mechanism behind our results could be that self-esteem changes preferences themselves, e.g. for socializing and experimentation. If this is true, then low self-esteem people may choose to live a more secluded lifestyle. They probably stay at home more and thus have more time at hand to focus on studying. Third, the low self-esteem individuals may feel that they will be disadvantaged later on in life, either in the sense that they need to work harder for equal outcomes or in the sense that they will be handicapped in certain domains such as social skills. Thus they may try to make up for this perceived handicap by putting more effort into academics.

Our findings add an important aspect to the existing literature: so far researchers have focused on the consequences of self-esteem for one single outcome. Life, however, is rarely uni-dimensional. At any given point in time there are usually many tasks at hand, and changes in self-evaluations may affect choice of effort and consequently performance in different tasks in different ways. In that sense our main result does in no way imply that increases in self-esteem always lead to lower performance in all domains. Quite the contrary: while it seems that high self-esteem individuals expend less academic effort, they perform well in other important areas, e.g. finding a job or socializing with other

people. While they accumulate less human capital, high self-esteem students may thus have comparative advantages in building other forms of economically relevant assets. This is merely a first pointer that the heterogeneity of the returns to self-esteem across tasks should be taken into account and lead to a more differentiated assessment of self-esteem – it should also be interesting to evaluate whether this is true for other non-cognitive skills as well.

6 Conclusion

Economists have recently gathered a considerable amount of evidence suggesting that positive self-views promote success in various circumstances of life. To the best of our knowledge, this paper is the first which makes use of a natural experiment to address the issue of whether positive self-views are inherently performance-enhancing. Using data from US high school students, we find that self-esteem is causally related to performance in our example, further supporting the view that psychology matters. However, the effect is opposite from what is usually postulated: Students with higher levels of self-esteem turn out to have a lower grade point average, i.e. perform worse in school. This effect is masked in standard OLS estimations, which throughout the various specifications yield the positive coefficients on self-esteem that are typically reported in the literature.

Our results add a new angle to the existing economics literature concerned with psychological factors by providing evidence for the potential downside of harboring self-esteem. This of course does not mean that self-esteem is always detrimental to all kinds of performance. In fact, we find that high self-esteem students who perform worse in school seem to spend more effort outside school, e.g. in socializing with other people, which suggests that the motivational effects of self-esteem are not uniform but rather heterogeneous across spheres. While exploring the determinants of school performance helps to understand the process of human capital formation and is therefore interesting in its own right, it remains left for future research to determine whether negative effects of self evaluations can be found in other economically relevant contexts, such as the workplace environment or other labor

market domains. It also should be interesting to see whether the upward bias we find in the self-esteem case carries over to other non-cognitive skills and personality traits. In general, the magnitude of the bias we found in our setup should encourage researchers to take into account possible endogeneity issues when dealing with psychological variables.

The results we have obtained also bear policy relevance. Enrolling juvenile offenders in self-esteem boosting programs in order to prevent recidivism remains a prevalent measure in the American juvenile justice system. In an example closer to our setup many schools offer self-esteem enhancing classes for their students where the message is perpetuated that a better evaluation of the self will lead to better educational performance as well as other desirable outcomes. At least in terms of school performance this does not seem to be the case and while it does not mean that boosting self-esteem is a bad measure per se, the aforementioned heterogeneity of effects should be taken into account: The target outcome may not even move in the intended direction and there might be side effects such that overall welfare effects are unclear. Taken together our results are a first pointer that standard policy recommendations may need to be reconsidered and that much more work is needed in order to fully understand how self-evaluations shape human behavior.

Acknowledgement

This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain the Add Health data files is available on the Add Health website (`http://www.cpc.unc.edu/addhealth`). No direct support was received from grant P01-HD31921 for this analysis.

References

Almlund, M., Duckworth, A. L., Heckman, J. J., and T. D. Kautz (2011), Personality Psychology and Economics, in: Hanushek, E., A., Machin, S. J., and L. Woessmann (eds.), Handbook of the Economics of Education, volume 4, chapter 1, 1-183, North-Holland.

Anderson, T. W., and H. Rubin (1949), Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations, *The Annals of Mathematical Statistics*, 20(1), 46-63.

Andrews D. W. K., and J. Stock (2005), Inference with Weak Instruments, *NBER technical working paper series* 313.

Angrist, J. D., and A. B. Krueger (1999), Empirical Strategies in Labor Economics, in: Ashenfelter, O., and D. Card (eds.), Handbook of Labor Economics, edition 1, volume 3, chapter 23, 1277-1366, Elsevier.

Angrist, J. D., and J.-S. Pischke (2009), Mostly Harmless Econometrics: An Empiricist's Companion, Princeton University Press.

Baumeister, R. F., Campbell, J. D., Krueger, J. I., and K. D. Vohs (2003), Does High Self-Esteem Cause Better Performance, Interpersonal Success, Happiness, or Healthier Lifestyles? *Psychological Science in the Public Interest*, 4 (1), 1-44.

Bataille, V., Snieder, H., MacGregor, A. J., Sasieni, P., and T. D. Spector (2002), The Influence of Genetics and Environmental Factors in the Pathogenesis of Acne: A Twin Study of Acne in Women, *Journal of Investigative Dermatology*, 119, 1317-1322.

Bénabou, R., and J. Tirole (2002), Self-Confidence and Personal Motivation, *Quarterly Journal of Economics*, 117(3), 871-915.

Borghans, L., Duckworth, A. L., Heckman, J. J., and B. Ter Weel, (2008), The Economics and Psychology of Personality Traits, *Journal of Human Resources*, 43

(4), 972-1059.

Bowles, S., Gintis, H., and M. Osbourne (2001), The Determinants of Earnings: A Behavioral Approach, *Journal of Economic Literature*, 39, 1137-1176.

Compte, O., and A. Postlewaite (2004), Confidence-Enhanced Performance, *American Economic Review*, 94(5), 1536-1557.

Dalgard, F., Gieler, U., Holm, J. O., Bjertness, E., and S. Hauser (2008), Self-esteem and Body Satisfaction Among Late Adolescents with Acne: Results From a Population Survey, *Journal of the American Academy of Dermatology*, 59(5), 746-751.

Davidovici, B. B., and R. Wolf (2010), The Role of Diet in Acne: Facts and Controversies, *Clinics in Dermatology*, 28(1), 12-16.

Duckworth, A. L., Steen, T. A., and M. E. P. Seligman (2005), Positive Psychology in Clinical Practice, *Annual Review of Clinical Psychology*, 1, 629-651.

Dufour, J.-M., and M. Taamouti (2005), Projection-Based Statistical Inference in Linear Structural Models with Possibly Weak Instruments, *Econometrica*, 73(4), 1351-1365.

Drago, F. (2011), Self-esteem and Earnings, *Journal of Economic Psychology*, 32(3), 480-488.

Feinstein, L. (2000), The Relative Economic Importance of Academic, Psychological and Behavioural Attributes Developed in Childhood, unpublished, LSE CEP Discussion Paper and University of Sussex, Discussion Paper in Economics, No. 62.

Fortin, N. M. (2008), The Gender Wage Gap among Young Adults in the United States: The Importance of Money vs. People, *Journal of Human Resources*, 43(4), 884-918.

Fryer, R. G., and P. Torelli (2010), An Empirical Analysis of 'Acting White', *Journal of Public Economics*, 94(5-6), 380-396.

Fulton, J. E., Plewig, G., and A. M. Kligman (1969), Effect of Chocolate on Acne Vulgaris, *Journal of the American Medical Association*, 210(11), 2071-2074.

Green, J., and R. D. Sinclair (2001), Perceptions of Acne Vulgaris in Final Year Medical Student Written Examination Answers, *Australasian Journal of Dermatology*, 42(2), 98-101.

Greenberg, J., Solomon, S., Pyszczynski, T., Rosenblatt, A., Burling, J., Lyon, D., Simon, L., and E. Pinel (1992), Why Do People Need Self-Esteem? Converging Evidence That Self-Esteem Serves an Anxiety-Buffering Function, *Journal of Personality and Social Psychology*, 63(6), 913-922.

Goldsmith, A. H., Veum, J. R. and W. Darity (1997), The Impact of Psychological and Human Capital on Wages, *Economic Inquiry*, 35(4), 815-829.

Halvorsen, J. A., Braae, O., Thoresen, M., Holm, J. O., and E. Bjertness (2008), Comparison of Self-reported Skin Complaints with Objective Skin Signs Among Adolescents,*Acta Dermato-Venereologica*, 88 (6), 573-577.

Heckman, J. J., Stixrud, J., and S. Urzua (2006), The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior, *Journal of Labor Economics*, 24(3), 411-482.

Heineck, G., and S. Anger (2010), The Returns to Cognitive Abilities and Personality Traits in Germany, *Labour Economics*, 17(3), 535-546.

James, W. D. (2005), Acne, *The New England Journal of Medicine*, 352, 1463-1472.

Judge, T. A., and J. E. Bono (2001), Relationship of Core Self-Evaluations Traits - Self-Esteem, Generalized Self-Efficacy, Locus of Control, and Emotional Stability - With Job Satisfaction and Job Performance: A Meta-Analysis, *Journal of Applied Psychology*, 86, 80-92.

Judge, T. A., Locke, E. A., and C. C. Durham (1997), The Dispositional Causes of Job Satisfaction: A Core Evaluations Approach, *Research in Organizational Behavior*, 19(1), 151-188.

Kurokawa, I., Danby, F. W., Ju, Q., Wang, X., Xiang, L. F., Xia, L., Chen, W. C., Nagy, I., Picardo, M., Schagen, S., Tsatsou, F., and C. C. Zouboulis (2009), New Developments in Our Understanding of Acne Pathogenesis and Treatment, *Experimental Dermatology*, 18(10), 821-832.

Lim, Y. L., Chan, Y. H., Yosipovitch, G., and M. W. Greaves (2008), Pruritus is a Common and Significant Symptom of Acne, *Journal of the European Academy of Dermatology and Venereology*, 22(11), 1332-1336.

Mallon, E., Newton, J. N., Klassen, A., Stewart-Brown, S. L., Ryan, T. J., and A. Y. Finlay (1999), The Quality of Life in Acne: A Comparison with General Medical Conditions Using Generic Questionnaires, *British Journal of Dermatology*, 140(4), 672-676.

Magin, P., Pond, D., Smith, W., and A. Watson (2005), A Systematic Review of the Evidence for 'Myths and Misconceptions' in Acne Management: Diet, Face-Washing and Sunlight, *Family Practice*, 22(1), 62-70.

Mueller, G., and E. Plug (2006), Estimating the Effect of Personality on Male and Female Earnings, *Industrial and Labor Relations Review*, 60(1), 3-22.

Nyhus, E. K., and E. Pons (2005), The Effects of Personality on Earnings, *Journal of Economic Psychology*, 26(3), 363-384.

Osborne Groves, M. (2005), How Important is Your Personality? Labor Market Returns to Personality for Women in the US and UK, *Journal of Economic Psychology*, 26(6), 827-841.

Passarini, B., Infusino, S. D., and E. Kasapi (2010), Chloracne: Still Cause for Concern, *Dermatology*, 221(1), 63-70.

Putnam, R. D. (2001), Bowling Alone. New York.

Reich, A., Trybucka, K., Tracinska, A., Samotij, D., Jasiuk, B., Srama, M., and J. C. Szepietowski (2008), Acne Itch: Do Acne Patients Suffer From Itching?, *Acta Dermato-Venereologica*, 88(1), 38-42.

Rosenberg, M. (1965), Society and the Adolescent Self-Image. Princeton, NJ: Princeton University Press.

Salerno, S. (2005), SHAM: How the Self-help Movement Made America Helpless. New York: Crown Publishers.

Tan, H. H., Tan, H. W., Barkham, T., Yan, X. Y., and M. Zhu (2007), Community-Based Study of Acne Vulgaris in Adolescents in Singapore, *British Journal of Dermatology*, 157(3), 547-551.

Thomas, D. R. (2005), Psychosocial Effects of Acne, *Journal of Cutaneous Medicine and Surgery: Incorporating Medical and Surgical Dermatology*, 8(4), 3-5.

Waddell, G. R. (2006), Labor-Market Consequences Of Poor Attitude And Low Self-Esteem In Youth, *Economic Inquiry*, 44(1), 69-97.

Webster, G. F. (2001), Acne Vulgaris and Rosacea: Evaluation and Management, *Clinical Cornerstone*, 4 (1), 15-22.

Williams, H. C., Dellavalle, R. P., and S. Garner (2012), Acne Vulgaris, *The Lancet*, 379(9813), 361-372.

Zouboulis, C. C., Eady, A., Philpott, M., Goldsmith, L. A., Orfanos, C., Cunliffe, W. C., and R. Rosenfeld (2005), What is the Pathogenesis of Acne?, *Experimental Dermatology*, 14(2), 143.

Table 7: Description of Variables.

Variable	Description
Main variables	
self-esteem	Rosenberg index of self-esteem, sums up the answers to the following 6 questions, all of which are scaled: disagree (1) - agree (5). Do you agree or disagree that you have many good qualities? Do you agree or disagree that you have a lot to be proud of? Do you agree or disagree that you like yourself just the way you are? Do you agree or disagree that you feel you are doing things just about right? Do you agree or disagree you feel socially accepted? Do you agree or disagree you feel loved and wanted?
acne	in the past 12 months, how often have you had skin problems, such as itching or pimples? never, a few times, more than weekly.
high school GPA	grade point average from HS transcripts. all subjects, year of survey. failed is coded as 0, so the range of this variable is 0 to 4, where higher numbers mean better grades.
Personal characteristics	
male	indicator variable, 1 if male.
health	How is your general health? Scale: excellent (1)- poor (6).
race	white, black, american indian, asian, other.
peabody intelligence test	verbal intelligence test, age adjusted.
HH income	total household income in 1000 US-$.
single HH	respondent lives with a single parent.
welfare	parent(s) receive(s) welfare payments.
parental education	highest degree attained by either parent: 'less than high school' to 'beyond college'.
age	age in years.
grade	respondent is in this high school grade.
bmi	body mass index. weight in kg divided by height in meters squared.
no medical treatment	respondent didn't seek medical care when he thought he should.
physical maturity	how advanced is your physical development compared to others your age? self assessed, look younger than most (1) - look older than most (5).
trouble teacher	in the past year how often did you have trouble getting along with your teachers? never (0) - every day (4).
trouble students	in the past year how often did you have trouble getting along with other students? never (0) - every day (4).
unfair teacher	the teachers at your school treat students fairly: strongly agree (1)- strongly disagree (5).
pill	indicator, do you currently take a birth control pill. zero for male respondents.
snack food	indicator, do you usually eat snack foods for breakfast. Wave I only.
chips	yesterday, did you eat potato chips, corn chips, tortilla chips, pretzels, or popcorn? Wave II only.
important brilliant student	indicator, parents view 'brilliant student' as the most important quality for her child. Wave I only.
Falsification variables	
hair under arms	how much hair is under your arms now? no hair (1) - a lot of thick hair (5).
lower voice	is your voice lower now than in grade school? same as in grade school (1) - a whole lot lower (5).
facial hair	how thick is the hair on your face? scattered/no hair (1) - like on a man's face (5).
ever menstruated	have you ever had a menstrual period or menstruated?
Channels	
drive mls	about how many miles do you drive each week? don't drive (1) to $>$ 100 mls (4). missing if age$<$16.
work hrs	how many hours do you spend working for pay in a typical non-summer week?
intercourse	have you ever had intercourse?
marijuana	respondent has ever tried marijuana.

Table 8: Summary statistics, Wave I.

	Mean	Std. Dev.	Min	Max	*N*
Main variables					
self-esteem	24.563	3.548	6	30	4927
acne	1.131	0.723	0	2	4927
high school GPA	2.586	0.912	0	4	4927
Personal characteristics					
male	0.49	0.5	0	1	4927
health	2.076	0.89	1	5	4927
white	0.686	0.464	0	1	4927
black	0.184	0.388	0	1	4927
am indian	0.011	0.106	0	1	4927
asian	0.064	0.244	0	1	4927
other	0.054	0.227	0	1	4927
peabody intelligence test	103.437	13.386	14	136	4927
HH income	48.942	46.082	0	870	4927
single HH	0.272	0.445	0	1	4927
welfare	0.082	0.274	0	1	4927
parental educ<HS	0.094	0.292	0	1	4927
parental educ=HS	0.274	0.446	0	1	4927
parental educ=some college	0.137	0.343	0	1	4927
parental educ=college grad	0.355	0.478	0	1	4927
parental educ>college	0.141	0.348	0	1	4927
age	16.121	1.112	13	21	4927
grade	2.24	0.994	1	5	4927
bmi	22.854	4.347	13.824	49.17	4927
no medical treatment	0.208	0.406	0	1	4927
res. looks younger than most	0.084	0.277	0	1	4927
res. looks younger than some	0.11	0.313	0	1	4927
res. looks about average	0.399	0.49	0	1	4927
res. looks older than some	0.278	0.448	0	1	4927
res. looks older than most	0.128	0.335	0	1	4927
trouble teacher	0.806	0.893	0	4	4927
trouble students	0.792	0.882	0	4	4927
unfair teacher	2.541	1.036	1	5	4927
pill	0.041	0.198	0	1	4927
snack food	0.071	0.257	0	1	4927
important brilliant student	0.683	0.466	0	1	4927
Falsification variables					
hair under arms	3.367	0.88	1	5	2406
lower voice	3.55	1.199	1	5	2411
facial hair	1.817	0.801	1	4	2407
ever menstruated	0.967	0.18	0	1	2509
Channels					
drive mls	2.177	0.986	1	4	3343
work hrs	8.131	11.249	0	120	4909
intercourse	0.416	0.493	0	1	4917
marijuana	0.314	0.464	0	1	4901

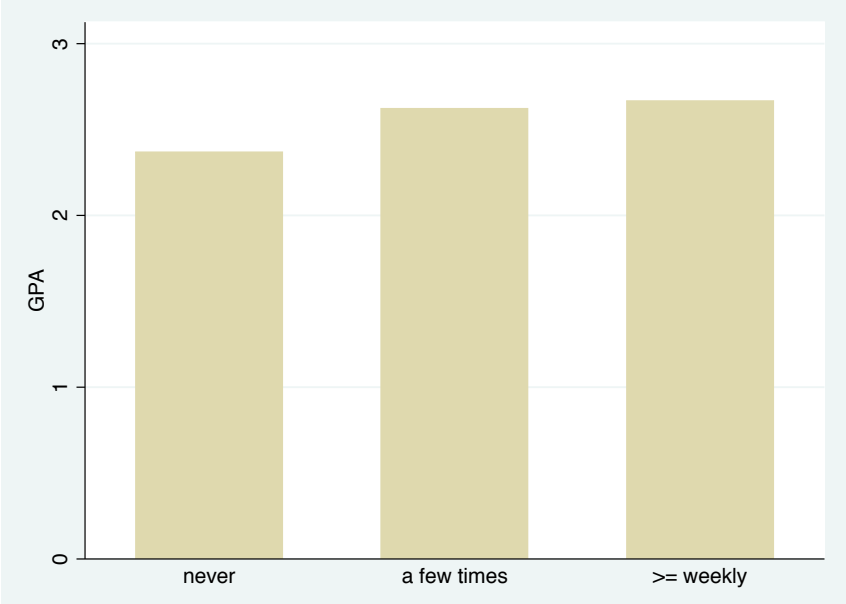
Note: Summary statistics are calculated for the sample used in the wave I baseline IV (see Table 2).

Table 9: Summary statistics, Wave II.

	Mean	Std. Dev.	Min	Max	*N*
Main variables					
self-esteem	25.136	3.38	7	30	3684
acne	1.102	0.715	0	2	3684
high school GPA	2.655	0.872	0	4	3684
lagged high school GPA	2.667	0.861	0	4	3684
Personal characteristics					
male	0.479	0.5	0	1	3684
health	2.031	0.857	1	5	3684
white	0.693	0.461	0	1	3684
black	0.181	0.385	0	1	3684
am indian	0.012	0.107	0	1	3684
asian	0.061	0.24	0	1	3684
other	0.053	0.223	0	1	3684
peabody intelligence test	103.689	13.58	14	131	3684
HH income	50.527	50.305	0	870	3684
single HH	0.239	0.427	0	1	3684
welfare	0.072	0.259	0	1	3684
parental educ<HS	0.084	0.278	0	1	3684
parental educ=HS	0.271	0.444	0	1	3684
parental educ=some college	0.131	0.337	0	1	3684
parental educ=college grad	0.365	0.481	0	1	3684
parental educ>college	0.15	0.357	0	1	3684
age	16.78	0.979	14	20	3684
grade	2.99	0.824	2	6	3684
bmi	23.221	4.646	13.574	50.791	3684
no medical treatment	0.2	0.4	0	1	3684
res. looks younger than most	0.091	0.288	0	1	3684
res. looks younger than some	0.115	0.319	0	1	3684
res. looks about average	0.41	0.492	0	1	3684
res. looks older than some	0.264	0.441	0	1	3684
res. looks older than most	0.119	0.324	0	1	3684
trouble teacher	0.706	0.830	0	4	3684
trouble students	0.738	0.86	0	4	3684
unfair teacher	2.585	1.009	1	5	3684
pill	0.053	0.223	0	1	3684
chips	0.491	0.5	0	1	3684

Note: Summary statistics are calculated for the sample used in the wave II baseline IV (see Table 4).

Figure 7: High school GPA and acne, Wave I.



Note: The figure displays the mean high school GPA by levels of acne. The sample is the same as the estimation sample used in Table 2. Number of observations $N = 4927$.

Table 10: First stage estimates, Wave I.

	IV first stage	
acne	$-.298^{***}$	(0.058)
younger than most[(a)]	$-.022$	(0.204)
younger than some	$-.163$	(0.159)
older than some	.167	(0.108)
older than most	$.515^{***}$	(0.159)
no medical treatment	$-.642^{***}$	(0.101)
pill	.208	(0.307)
BMI	$-.0082$	(0.012)
snack food	.118	(0.153)
important brilliant student	$-.0885$	(0.126)
trouble w/ teacher	$-.165^{***}$	(0.061)
trouble w/ students	$-.343^{***}$	(0.061)
teachers unfair	$-.439^{***}$	(0.055)
basic individual characteristics		
male	$.95^{***}$	(0.111)
health very good[(b)]	-1.23^{***}	(0.118)
health good	-2.00^{***}	(0.133)
health fair	-2.94^{***}	(0.234)
health poor	-5.19^{***}	(1.000)
black[(c)]	$.899^{***}$	(0.232)
am. indian	.174	(0.507)
asian	$-.556^{**}$	(0.265)
other	$-.339$	(0.247)
ability/family controls		
peabody intelligence	$-.0029$	(0.004)
HH income	$7.6e-04$	(0.001)
single HH	$-.0822$	(0.142)
welfare	$6.9e-04$	(0.180)
parents high school[(d)]	.0383	(0.219)
parents some college	.0358	(0.218)
parents college grad	.215	(0.236)
parents > college	.105	(0.263)
age fixed effects	*yes*	
grade fixed effects	*yes*	
grade*age fixed effects	*yes*	
school fixed effects	*yes*	
observations	4927	

Note: Standard errors in parentheses allow for clustering at the school level. First stage estimates for Wave I. Reference categories are: (a) looks neither younger nor older, (b) health excellent, (c) white, (d) parents' education less than high school. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 11: Reduced form estimates, Wave I.

acne	.0602***	(0.015)
younger than most[(a)]	−.116**	(0.053)
younger than some	−.0418	(0.061)
older than some	−.129***	(0.032)
older than most	−.117***	(0.039)
no medical treatment	.0073	(0.028)
pill	−.0524	(0.053)
BMI	−.0076**	(0.004)
snack food	−.0896*	(0.046)
important brilliant student	.0113	(0.022)
trouble w/ teacher	−.167***	(0.019)
trouble w/ students	−.0356***	(0.013)
teachers unfair	−.0692***	(0.013)
basic individual characteristics		
male	−.303***	(0.023)
health very good[(b)]	−.0764***	(0.024)
health good	−.183***	(0.033)
health fair	−.264***	(0.056)
health poor	−.39*	(0.219)
black[(c)]	−.118***	(0.042)
am. indian	−.105	(0.120)
asian	.355***	(0.036)
other	.0294	(0.073)
ability/family controls		
peabody intelligence	.0169***	(0.001)
HH income	8.3e−04***	(0.000)
single HH	−.093***	(0.023)
welfare	−.0412	(0.041)
parents high school[(d)]	.045	(0.043)
parents some college	.138***	(0.047)
parents college grad	.233***	(0.050)
parents > college	.358***	(0.052)
age fixed effects	*yes*	
grade fixed effects	*yes*	
grade*age fixed effects	*yes*	
school fixed effects	*yes*	
observations	4927	

Note: Standard errors in parentheses allow for clustering at the school level. First stage estimates for Wave I. Reference categories are: (a) looks neither younger nor older, (b) health excellent, (c) white, (d) parents' education less than high school. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.