Discussion Paper No. 71
Williams Project on the Economics of Higher Education
23 Whitman Street
Mears West
Williams College
Williamstown, MA 01267

http://www.williams.edu/wpehe

A Note on How Well Available Income Information Identifies Low-Income Students

Gordon Winston, *Williams College*
Catharine Hill, *Vassar College*
David Zimmerman, *Williams College*

March, 2007
DP-71

Acknowledgements: Our debt to the Andrew W. Mellon Foundation for its support of the Williams Project on the Economics of Higher Education is deep and continuing. Bill Bowen encouraged this investigation at an early stage and we relied heavily on our knowledgeable colleagues at Williams including Paul Boyer, Michelle Waryjasz, Chris Winters, Dick Nesbitt, Connie Sheehy, Mark Robertson, Henry Bruton, Ralph Bradburd and John Chandler. Carey Aubert provided thoughtful and essential help with numbers, tables, graphs and organization.

ABSTRACT

This note looks at the quality of the information on family income that selective colleges rely on to increase equality of opportunity by recruiting high-ability, low-income students. Individual family income estimates embedded in the College Board's search parameters are compared, for 635 recent Williams matriculants, with their incomes as reported on IRS Forms 1040 and, for further comparison, with self-reported incomes. The data suggest that there is considerable room for improvement and, indeed, until there is better information, that any effort to increase equality of opportunity by energetic recruitment of high-ability, low-income students will be haphazard at best.

With sharply increased attention to the admission of high-ability low-income students at the nation's most selective colleges, it has become important to be able to identify those students in order to bring them into the applicant pool and monitor their progress through the admission process and beyond. Conventional searches of student test-takers identify those whose scores suggest high ability, but information on family incomes rests on less reliable measures.

So this note focuses on the primary source of information on family incomes of applicants to selective colleges and universities: College Board Neighborhood Cluster income estimates and, for comparison, student self-reported incomes. We have a small but appropriate population of students for whom there is information on family incomes from tax forms 1040 as well as one or the other source of family income estimate. Presuming the 1040 is as close to the truth about family income as we're likely to get, each income estimating method can be evaluated relative to that truth.

The population consists of the 635 students who entered Williams College in 2005 and 2006 and also submitted 1040 income information in applying for financial aid. Of these, 198 came through the College Board search process so were also assigned a Neighborhood Cluster estimate of family income. Another (overlapping) 400 students provided self-reported estimates of family income when filling out Williams' Admitted Student Questionnaire (ASQ). For these two methods, the individuals' income estimates were then compared to their 1040 family incomes. Note, importantly, that this approach escapes the tyranny of averages: evaluation is based on the comparison, for each individual, of his or her own actual income with his or her estimated family income rather than on average values for a group.

This population is not ideal for an evaluation of the *general* accuracy of either Neighborhood Cluster or self-reported income estimates[1] since these are low-income students who applied to Williams, were admitted, and chose to attend. They can, therefore, be taken as representative of matriculated students at a highly selective private college but not of low-income high-ability students in the general population: too many selective steps have intervened – a high-ability, low-income student from the general population has to *know* of Williams, *apply*, be *admitted*, and, finally, *matriculate* and each of those steps might introduce serious selection effects, making our results quite inapplicable to the general population of low-income students. Of course, they might not. Ultimately, the question at issue is the accuracy of income estimates for the highly selective colleges like Williams that are aggressively seeking to enroll more low-income, high-ability students.

The data here are used to address four questions: (1) of those students whose family incomes are actually (1040s) in the bottom quintile (or two) of the national population,[2] what proportion is identified by each of these income estimates? Going in the other direction, (2) of those who are assigned by one of these estimating methods to the bottom income quintile (or two), what proportion actually belong there? (3) Of those who are misclassified, how big is the mistake? And, finally, (4) given both the accuracy and availability of each estimating method, what proportion of the low-income students does it identify?

[1] Evidence on the general accuracy of self-reported incomes, more generally, is summarized in the appendix to Winston and Hill, "How Scarce Are High-Ability, Low-Income Students?" in Michael S. McPherson and Morton Owen Schapiro, Editors, *College Access: Opportunity or Privilege?* New York, The College Board, 2006.

[2] We use the national distribution of family incomes by quintile both to give the broader context in which to evaluate income and to tie it to our earlier work. In 2004, the quintile boundaries were: Quintile 1, $0 to $24,780, Quintile 2, $24,781 to $43,400, Quintile 3, $43,401 to $65,832, Quintile 4, $65,833 to $100,000, Quintile 5 above $100,000.

Neighborhood Cluster Income Estimates

Table 1 reports students' College Board Neighborhood Cluster income estimates (NC) and how they compare with actual incomes from families' 1040 tax returns. Keep in mind that of the 635 students who provided 1040 income information to Williams, we have Neighborhood Cluster income estimates for only 198 of them – those who came through the College Board student search process – so the initial comparisons are within that set for whom we have both kinds of income information and therefore our results are relevant to the accuracy, not the availability, of the income estimation method for this group of students.

Table 1

A. Neighborhood Cluster Income Estimates and Tax Return Incomes
(635 Matriculated Students, 2005 and 2006)

	Actual (1040) Family Incomes by U.S. Census Quintiles					
Neighborhood Cluster (NC) Income Estimates	**Q-1**	**Q-2**	**Q-3**	**Q-4**	**Q-5**	**Total**
No NC Estimate	80	51	54	72	180	**437**
NC-1	1	0	0	0	0	**1**
NC-2	8	9	9	13	13	**52**
NC-3	7	5	11	14	35	**72**
NC-4	3	11	8	13	33	**68**
NC-5	0	1	0	2	2	**5**
Number with NC Estimates	19	26	28	42	83	**198**
Total	**99**	**77**	**82**	**114**	**263**	**635**

.

B. The Distribution of Low-Income Students over NC Income Estimates

Q-1 students with NC Estimates	Q-1 Identified as such by NC	Q-1 Identified as Q-2 by NC	Q-1 Identified as Q-3 by NC	Q-1 Identified as Q-4 by NC	Q-1 Identified as Q-5 by NC	Total
	1	8	7	3	0	19
% of Actual Q-1 Students	5%	42%	37%	16%	0%	100%

Q-1 + Q-2 Students with NC Estimates	Q-1 + Q-2 Identified as such by NC	Q-1 + Q-2 Identified as Q-3 by NC	Q-1 + Q-2 Identified as Q-4 by NC	Q-1 + Q-2 Identified as Q-5 by NC	Total
	18	12	14	1	**45**
% of Actual Q-1+Q-2 Students	40%	27%	31%	2%	100%

C. The Distribution of NC Low Income Estimates over Actual Student Incomes

	Actual (1040) Family Incomes by U.S. Census Quintiles					
Students Classified as NC-1	**Q-1**	**Q-2**	**Q-3**	**Q-4**	**Q-5**	**Total**
	1	0	0	0	0	1
% of NC-1	100%	0%	0%	0%	0%	100%

NC Classified NC-1 or NC-2	Actually in Q-1 or Q-2	Actually in Q-3	Actually in Q-4	Actually in Q-5	Total
	18	9	13	13	53
% of NC-1 and NC-2	34%	17%	25%	25%	100%

Table 1A shows that 19 of these 198 students come from families whose actual 1040 incomes are in the first quintile (under $24,780 in 2004), but that the Neighborhood Cluster income estimates put only one of them in that bottom quintile ("NC-1"). So, charitably, 5% of the poorest (Quintile One) students are correctly identified in these NC data. The remaining lowest-income students are mis-identified as being in Quintile Two (8 of them or 42%), Quintile Three (7 or 37%) and Quintile Four (3 or 16%). Looking at the bottom 40% of the income distribution – Quintiles One and Two, together – there are 45 students in those two quintiles by 1040 information and of those, 18 or 40% are so identified by the NC measure (NC-1 or NC-2). Twelve (27%) are put in Quintile Three by the Neighborhood Cluster measure (NC-3), fourteen in NC-4 (31%) and one (2%) in NC-5.

Viewed from the other direction, Table 1C describes the shape and nature of the mistakes being made by the Neighborhood Cluster estimates. The one student put by Neighborhood Cluster in Quintile One ("NC-1"), belongs there, but considering the 53 students whom the NC income estimates put in the bottom two quintiles (NC-1 and NC-2 together), 18 of them belonged there (34%), 9 belonged in NC-3 (17%), 13 in NC-4 (25%), and 13 in NC-5 (25%). So, many of those whom Neighborhood Clusters identified as low-income students were in fact far from it – fully a quarter of those described as being in the bottom 40% of the nation's family income distribution were in the top 20% while 25% of those described as low-income were, in fact, in the fourth quintile and 17% in the third – fully half were in the top 40%.

Students' Self-Reported Incomes

Table 2 describes students' self-reported family incomes (on Williams' Admitted Student Questionnaire) and, again, the incomes reported for those students on tax Forms 1040. Here, 401 of the 635 students who submitted 1040 information also provided a self-reported income estimate on the ASQ so that set is the basis of comparison. Of those 401 students, 73 are (by 1040 information) in Quintile One and their self-reported incomes accurately put 52 of them there (71%). Self-reports put 14 of them (19%) in Quintile Two, 3 (4%) in Quintile Three, none in Quintile Four and 4 in Quintile Five (5%). Looking at the bottom two quintiles, together, in Table 2B, 125 students' 1040 incomes put them in Quintiles One or Two while their self-reported incomes put 110 of them there (ASQ-1 or ASQ-2) – that's 88%. Ten (8%) were put in Quintile Three, none in Quintile Four, and five (4%) in Quintile Five.

Table 2

A. Students' Self-Reported Incomes (ASQ) and Tax Return Incomes
(635 Matriculated Students, 2005 and 2006)

	Actual (1040) Family Incomes by U.S. Census Quintiles					
ASQ Self-Reported Incomes	**Q-1**	**Q-2**	**Q-3**	**Q-4**	**Q-5**	**Total**
No Report	26	25	30	49	104	**234**
ASQ-1	52	19	8	4	2	**85**
ASQ-2	14	25	16	3	4	**62**
ASQ-3	3	7	19	21	11	**61**
ASQ-4	0	0	7	35	24	**66**
ASQ-5	4	1	2	2	118	**127**
Number Reported	73	52	52	65	159	**401**
Total	**99**	**77**	**82**	**114**	**263**	**635**

B. The Distribution of Low-Income Students over ASQ Self-Reports

Q-1 Students with ASQ Self-Reports	Q-1 Identified as Such by Self-Report	Q-1 Identified as Q-2 by Self-Report	Q-1 Identified as Q-3 by Self-Report	Q-1 Identified as Q-4 by Self-Report	Q-1 Identified as Q-5 by Self-Report	Total
	52	14	3	0	4	**73**
% of Q-1 Students	71%	19%	4%	0%	5%	100%

Q-1 + Q-2 Students with ASQ Self-Reports	Q-1 + Q-2 Identified as such by Self-Report	Q-1 + Q-2 Identified as Q-3 by Self-Report	Q-1 + Q-2 Identified as Q-4 by Self-Report	Q-1 + Q-2 Identified as Q-5 by Self-Report	
	110	10	0	5	**125**
% of Q-1 + Q-2 Students	88%	8%	0%	4%	100%

C. The Distribution of Self-Reports over Actual Student Incomes

	Actual (1040) Family Incomes by U.S. Census Quintiles					
Students Reporting ASQ-1	**Q-1**	**Q-2**	**Q-3**	**Q-4**	**Q-5**	**Total**
	52	19	8	4	2	**85**
% of ASQ-1	61%	22%	9%	5%	2%	100%

Students Reporting ASQ-1 or ASQ-2	Actually in Q-1 or Q-2	Actually in Q-3	Actually in Q-4	Actually in Q-5	
	110	24	7	6	**147**
% of ASQ-1 + ASQ-2	75%	16%	5%	4%	100%

From the other direction, Table 2C shows that of the 85 students who self-reported incomes in the first quintile (ASQ-1), 52 of them (61%) were correct, 19 had actual incomes that put them in the second quintile (22%), 8 belonged in the third quintile (9%), 4 in the fourth (5%) and 2 in the fifth (2%). Looking again at the 147 students in the bottom two quintiles, 110 (75%) belong there, 24 (16%) belong in Quintile Three, 7 in Quintile Four (5%) and six (4%) in Quintile Five.

Figure 1A

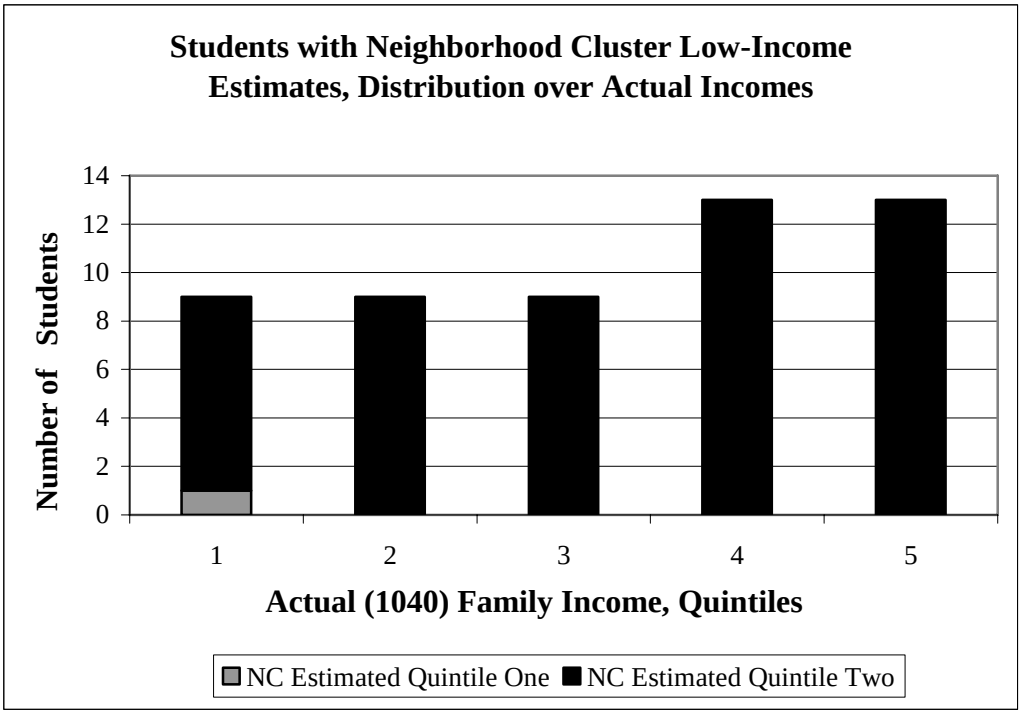
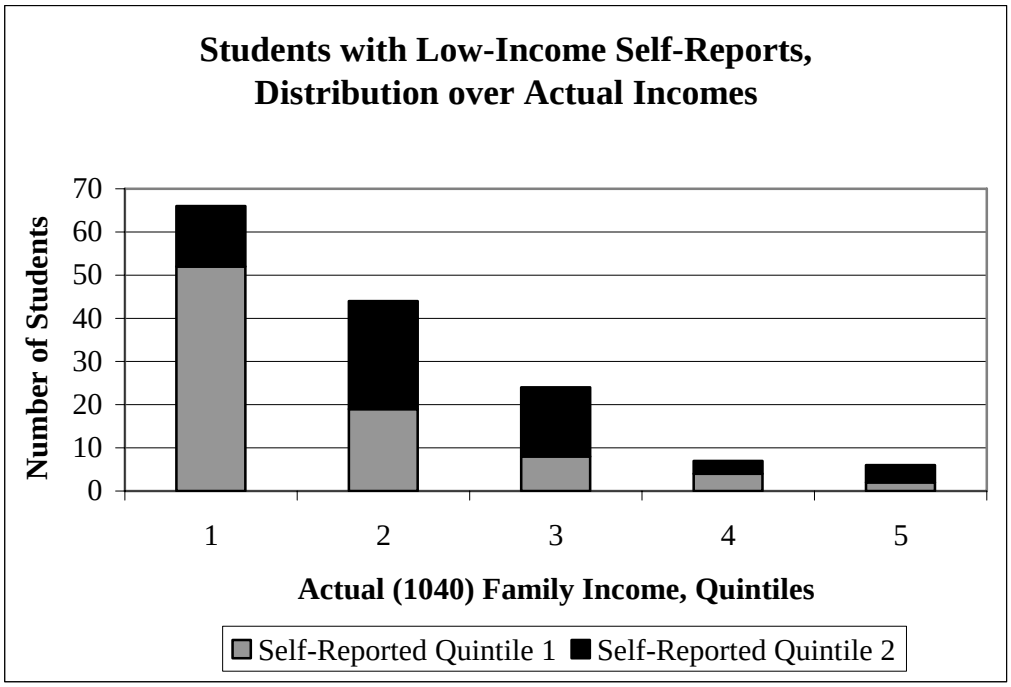
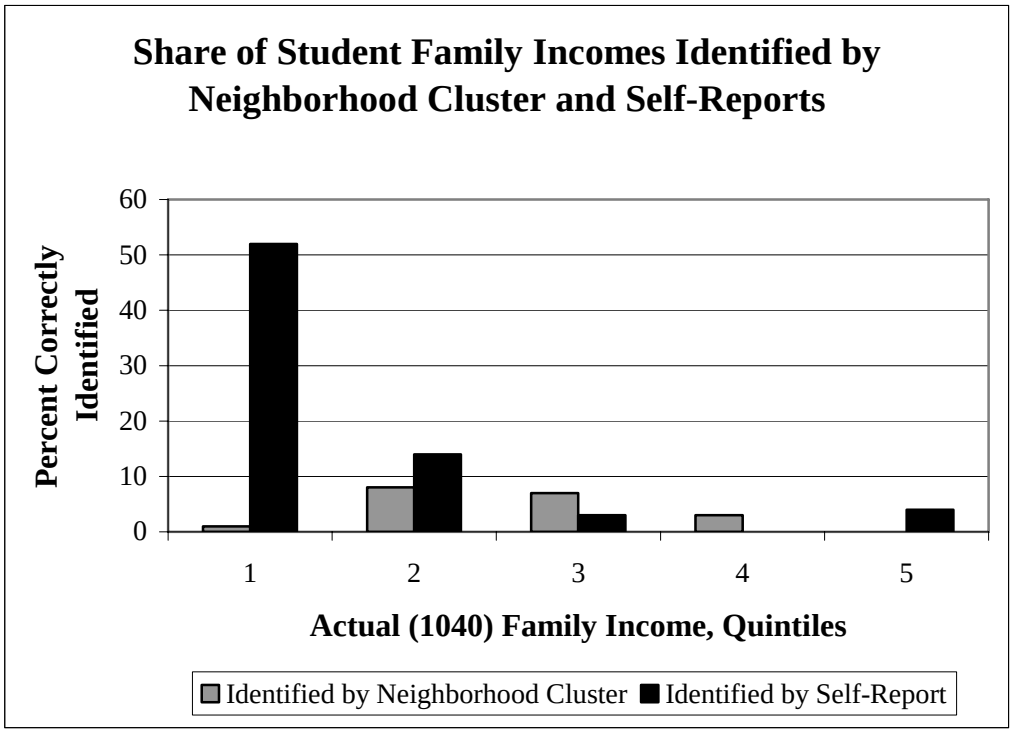


Figure 1B



Figures 1A and 1B summarize what we've seen. The height of the bars indicates the distribution over the 1040 income quintiles of those students estimated to fall into Quintile One (the bottom 20%) or Quintiles One and Two (the bottom 40%). Figure 1A describes the distribution of Neighborhood Cluster income estimates from Table 1A data and Figure 1B the distribution of self-reported income estimates from Table 2A. So one of the 19 students identified as being in Quintile One by Neighborhood Cluster actually is in Quintile One on the evidence of tax returns and 18 of the 45 identified by NC as coming from the bottom 40% of the US income distribution actually have incomes that put them there (Table 1). Figure 1B shows the same results for the Self-Reported incomes (Table 2**).**

Figure 2



A more demanding test of the usefulness of these income estimates is shown in Figure 2. It simply describes the share of the low-income students in a population like this, by quintile, that each method of income estimation identifies, given both the accuracy of the estimate – described above – and the availability of that kind of income estimate in the population. So, other things equal, the method that was available for a larger number of students might fare better regardless of its accuracy in the terms just described. Or vice versa. Earlier, we asked how well the estimating method fared when we had an income estimate and 1040 'truth' for everyone; here we ask how useful an estimate is when we have only the incomplete set of estimates we've actually got.

That said, against the 99 students actually in Quintile One among the 635 students, the NC estimates accurately identify only one of them as noted above. Considering Quintiles One and Two together, there are 176 such students in the population of 635 and NC identifies 18 or 10% of them. Self-reported incomes do a good deal better, identifying 43 of the 99 students in Quintile One (43%) and 47% of those in the bottom two quintiles; these estimates are better both because self-reports are more accurate and because we have self-reported income information on more students.

Conclusion

Family income information has come to serve, broadly, two purposes in the admissions process for selective schools: bringing low-income, high-ability students into the applicant pool in service of equality of opportunity – identifying and recruiting them – and monitoring their progress through the admission (and financial aid) process. Search and evaluation.

A primary strength of the Neighborhood Cluster income estimates has been their availability to guide student search in time for a college to initiate contact and encourage application. So even if these income estimates are not very accurate, they might be the best available in time to be used for the search and recruitment of low-income students.

In the admissions process itself, any effort to pay special attention to high-ability low-income students – including "putting a thumb on the scale"– clearly needs to know who they are. The most straightforward way to make more accurate information on low-income families

available for admissions evaluation, then, might well be simply to invite students voluntarily to submit a copy of the family's 1040 with their application under a guarantee that its information would serve only to benefit the student's chances of admission or leave them unaffected. The school would guarantee, in other words, not to use income information to bias selection toward the full-pay student and against those who need significant aid.

Just how serious the income information problem is, we think, is suggested by the results in this paper. It is quite difficult for colleges to identify low-income students to search them out and encourage their application. We're fine with information on test scores but not with that on family incomes so identification of "high-ability, low-income students" is inherently haphazard. We hope that the results presented here may help motivate efforts to improve income estimates.