DEVELOPING AGRICULTURAL SECTOR: FROM QUALITY OF AGRICULTURAL EDUCATION TO EVALUATION OF POLICIES ENHANCING AGRICULTURAL VALUE CHAINS

A Dissertation

by

ANNA YERITSYAN

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Committee Chair,	James W. Mjelde
Committee Co-Chair,	Ariun Ishdorj
Committee Members,	Marco Palma
	Ray Charles "Chuck" Howard
Head of Department,	Rodolfo Nayga

August 2023

Major Subject: Agricultural Economics

ABSTRACT

Agriculture may not be a field that young people perceive as “fashionable” and find attractive as a future specialization. The importance of equipping world economies with food production specialists, however, cannot be overstated. Similarly, it is crucial to put together the right policies to ensure enabling environment for these specialists to work and supply the global population with food. To address these issues, the overall objective of this dissertation is enhancing agricultural value chains through revealing factors affecting quality of agricultural education and highlighting gaps in policies that if addressed may increase the impact of interventions at different stages of agricultural value chains. The first two essays of this dissertation use mixed effect hierarchical models and utilize data from Texas A&M University College of Agriculture and Life Sciences (COALS) to discuss the importance of universities’ grading standards and to reveal factors affecting the quality of education. The third essay examines knowledge gaps relating to the policies aimed at agricultural value chains development through application of the Evidence Gap Map (EGM) method.

First essay discusses the notion of grade inflation and examines factors that may impact grades. The main finding is systematic increase in grades from 1985 to 2019 after controlling for the influence of institution, instructors, and student factors. This finding suggests that grading may not be an accurate signal for evaluating quality of education.

The second essay reveals differences in grading patterns among departments in COALS and even within the same department between time periods. It appears differences in grades are mainly driven by specifics of each department. Because of these differences, it is important to exercise caution when comparing students and their grades across different majors, as this can lead to misleading conclusions for employers and graduate school recruiters.

In the third essay, policies that may impact agricultural productivity, market access, and farmers' welfare are identified. Based on the data from one hundred ninety-three studies, most frequently occurring policy interventions are related to input supply stage. The least frequently occurring interventions are related to governance along Agricultural Value Chains (AVC) and interventions targeted toward gender, poverty, and social issues.

ACKNOWLEDGEMENTS

I would like to express my gratitude to my committee chair Dr. Mjelde, for his guidance, coaching, and encouragement throughout the course of this research and during my PhD studies.

I am thankful for support I had from my committee co-chair Dr. Ishdorj, as well as Norman Borlaug Institute for International Agriculture staff, especially from Dr. Norton and Mrs. Tabares while guiding and supervising my work on agricultural value chains.

I really appreciate the support of my committee members Dr. Palma and Dr. Howard.

Special thanks to Dr. John Nichols, who encouraged me to start my PhD and guided throughout as a good friend and a wonderful mentor.

Appreciations also go to my friends and classmates and TAMU Agricultural Economics department faculty and staff, for making my time at Texas A&M University a great experience.

Finally, thanks to my family for their love and support.

CONTRIBUTORS AND FUNDING SOURCES

Author would like to thank Elizabeth Tabares, Prof. Margaret Foster, Qinye Jiang, Mengqiao Liu, Vikas Mishra, Thuy Nguyen, Siwan Song and Yaqing Zheng for their help while screening and extracting data for the third essay of this dissertation.

Funding Sources

The third essay of this dissertation was funded by the USDA Food for Progress Project FX18TA-10960R029 and was supported by Norman Borlaug Institute for International Agriculture at Texas A&M University.

TABLE OF CONTENTS

LIST OF TABLES

Page

INTRODUCTION

Designing policies to improve education and knowledge for the agricultural sector (beginning with the training and preparation of a skilled workforce and extending to all elements of the supply chain from inputs to global trade and retail) is crucial for global food security, especially with a growing world population. A wide range of policies and interventions (actions, programs, or activities) have been implemented in the agricultural sector. These policies start with the education of the future agricultural workforce and extend into agricultural production and beyond subsectors. Grading impact the preparedness of the future workforce. Policies that facilitate access to inputs, credit, and insurance help farmers mitigate risks and enable farm growth and efficient operations are other examples of policies impacting agriculture. Further policies facilitating the development of infrastructures such as roads, storage facilities, and logistic centers help connect farmers with markets and increase profits in the agricultural sector. Within this setting, the overall objective of this dissertation is enhancing agricultural value chains through revealing factors affecting quality of agricultural education and highlighting gaps in policies that if addressed may increase the impact of interventions at different stages of agricultural value chains.

To achieve the overall objective, this dissertation examines three different aspects of education and knowledge creation in the agricultural sector. The first essay, Chapter II, raises concerns about grade improvement versus grade inflation. Grade inflation refers to the phenomenon where grades are rising over time, without a corresponding improvement in the quality of students' work (Bejar and Blew, 1981). In contrast, grade increase occurs when students are earning higher grades because of improved learning. Specifically, three interrelated objectives are to examine: 1) has grade inflation occurred, 2) has the distribution of grades as measured by the standard deviation changed; and 3) how various factors influence class mean GPA and grade distributions. To meet these objectives a unique data set of instructor, institutional, and student characteristics is used to estimate if grade inflation is incurring in College of Agriculture and Life Sciences (COALS) at Texas A&M University. Data are from a variety of sources with the main source of grades, institutional, and student characteristics being Texas A&M University's registrar's office. Faculty information was obtained from university catalogs, internet searches, and contacting departments and instructors directly. Student grade information is highly protected by privacy rights laws. Because of this privacy, this research uses class level data rather than individual student data. This class level data contains information on 17,696 observations from Fall 1985 to Spring 2019. Previous literature suggests that differences in grading standards are observed not only between different universities (Popov and Bernhardt, 2013), but also between departments within the same university (Sabot and Wakerman-Linn, 1991; Herron and Markovich, 2017), and even between instructors in the same department (Jewell and McPherson, 2012). A three-stage hierarchical model is used to account for those differences. The results show that there was a systematic increase in mean class GPAs and standard deviation of grades in COALS. All three groups of characteristics are important in explaining grades.

The essay presented in Chapter II analyzes grades in COALS for all departments together. Because departments are found to potentially grade differently, the essay in Chapter III focuses on individual departments quantifying differences in grading patterns among the departments. The purpose of this is to determine whether grading patterns found in Chapter II are prevalent across all departments or if they are primarily driven by a small number of departments. The database from Chapter II is used to estimate two-stage hierarchical models for individual departments to enable comparisons across departments. Further the data is separated into two time periods to provide developments for each department across time. Specific objectives are to determine if potential grade inflation has been occurring by department and if it differs overtime, as well as examine factors influencing mean class GPA among different departments in COALS to provide information on factors causing these differences and explore if the influence has changed overtime. Examination of differences in grading patterns shows that potential grade inflation occurred in four departments in the first period (1989 – 2003) and in eight departments in the second period (2004 – 2019). Weak evidence exists that supports previous studies' claims that differences exist between STEM and non-STEM designated departments.

Chapter IV relates to policies beyond the formal education sector that shape and coordinate the agricultural sector. This chapter collects systematic evidence on agricultural policies and corresponding outcomes from low, middle, and upper-middle income countries from 2000 to 2022. Objectives are identifying extent and characteristics of existing empirical evidence regarding policy interventions that enhance AVC, as well as highlighting effectiveness of models of collaboration among local and international actors to support policy change.

Policy interventions are grouped according to stages of the value chain (input distribution, production, harvest, and post-harvest, processing, transportation and storage, marketing, and governance), while resulting outcomes are grouped by area of impact (production, market expansion, product movement in space and time, risk reduction, and improved social welfare related outcomes). A total of 7,021 studies from numerous databases, libraries and websites were screened, however only 193 satisfied the inclusion criteria and are selected for data extraction. All possible intervention / outcome pairs were mapped into Evidence Gap Map (Snilstveit et al., 2017) to show areas that were previously underexplored and may result in higher impact per unit of investment. These underexplored areas suggest further research topics for the researchers. The study also benefits policymakers. By highlighting specific types of interventions that can produce desired outcomes, it helps enable the effective allocation of limited resources. Further, if policymakers are constrained to certain types of interventions, it would help to understand what are the most likely outcomes that can be achieved through those interventions. For instance, although input supply stage of the value chain has the highest number of interventions, the highest percentage of studies discussing collaboration models was aimed at improving the markets stage, followed by the production and post-harvest stage of the value chain. This suggests that when faced with limited resources, AVC participants would rather improve market access, production, and post-harvest handling, rather than solely focus on researching the input supply stage which may already be well-explored.

GRADE INFLATION OR GRADE INCREASE

Introduction

Concerns about the upward trend in university grades have been voiced by academia and industry. The questions raised include "Are students better prepared or are higher grades being granted easier?" History of concerns about grade increases dates back to at least 1894, when the "Report of the Committee on Raising the Standard, Harvard University, 1894" (as cited in Kohn, 2002, p. 1) stated students are sometimes receiving A and B grades too readily. This upward trend is often noted as grade inflation, however, not all grade increases should be labelled inflation. Of the multiple definitions of grade inflation (Schutz et al., 2015), the definition adopted here is the one used by Kostal, Kuncel, and Sackett (2016 p. 13) which states "... increases in grades across time that do not reflect changes in the quality of students' work (Bejar and Blew, 1981)." The key component of this definition is not reflecting "changes in quality of student work."

Grade inflation raises concerns about the credibility of academia's standards of excellence and accountability through the lowering of academic standards (Kolevzon, 1981). Lowering standards through grade inflation raises the additional concern that grades are losing their ability to differentiate students' work (Kamber and Biggs, 2004; Pattison, Grodsky, and Muller, 2013; Peace, 2017). This inability to differentiate students' work arises because grades have a ceiling, one cannot give students a grade higher than the university's highest grade (Kamber and Biggs, 2004). Studies indicate grade inflation is not only a concern in the U.S. (Marini et al., 2018; Baird, Carter, and Roos, 2019), but there is interest in the phenomenon internationally (Bachan 2018).

Multiple forces have come together to produce grade inflation. Schutz et al. (2015) list 27 reasons for grade inflation that have been proposed in the literature. Within these forces, one finds evidence that instructor, university, and student characteristics may lead to grade inflation (Peace, 2017). Universities competing for reputation, competition for students, and increasing tuition and costs may lead universities to permit grade inflation to positively influence their reputation and justify tuition increases (Jewell, McPherson, and Tieslau, 2013). Instructors may inflate grades for reasons such as favorably biasing the student evaluations, pressure from their system to maintain student satisfaction, and job security (Eiszler, 2002; Kezim, Pariseau, and Quinn, 2005; Jaschik, 2008; Pease, 2017). Students place pressure on the system for high GPAs because of financial aid considerations and job placement. The notion that with increased tuition comes the right for high grades may also lead to grade inflation (Peace, 2017). Studies such as Boretz (2004) contradict these consumerism conclusions, suggesting students would rather earn higher grades. She notes the widespread acceptance of grade inflation is potentially damaging to universities. Although grades may be at an all-time high, Boretz (2004) argues faculty development programs and increased study support services have led to the increases. Along these lines, Jephcote, Medland, and Simon Lygo-Baker (2021) argue unexplained increases in grade should not automatically be viewed as grade inflation. Blindly labelling increases as grade inflation is damaging to higher education. Rather one should see opportunities to investigate the institutional context the grades are occurring.

Grade inflation is an empirical issue; factors, however, can give rise to higher grades that are not attributed to grade inflation as defined above. A positive shift in grade distribution that is caused by increases in achievement or learning is grade improvement, not grade inflation (Mostrom and Blumberg, 2012). Changing demographics and ability of students, increased student support services, improved instruction, and faculty development programs are some factors that can lead to grade improvement (Kuh and Hu, 1999; Boretz, 2004; Mostrom and Blumberg, 2012). Further, Boretz (2004) and Watjatrakul (2014) suggest that the move toward consumerism in higher education increases the quality of education through supporting student learning thus leading to higher grades. The idea of grade improvement was further developed by Flynn (1987), who came up with the phenomena known as The Flynn Effect. In his work, the results of IQ tests in schools are increasing on average approximately three points per decade. This finding is supported by Trahan et al. (2014) meta-analysis that finds grades for standardized tests are increasing, which may be attributed to an increase in the population's intelligence quotient.

The current study uses a unique data set of university, instructor, and student specific characteristics to examine grades in the College of Agriculture and Life Sciences (COALS) at Texas A&M University (TAMU) from 1985 to 2019. Specifically, three interrelated objectives are to examine: 1) has grade inflation occurred, 2) has the distribution of grades as measured by the standard deviation changed; and 3) how various factors influence class mean GPA and grade distributions.

This study's contributions include the use of a detailed data set to identify the effect of specific variables. Further, this study explicitly examines a college of agriculture at the class level instead of department or university level GPAs. No previous study was found with this explicit, detailed focus at a class level. COALS is one of the largest colleges of agriculture, awarding the most degrees in agriculture in the U.S. (DATAUSA, 2020) in 2019, approximately

3.5% of the degrees awarded nationally. Further, being a land grant university, COALS missions are similar to other land grant universities.

Literature Review

Given the importance of student success to both the student and university, it is not surprising the large number of theories and factors proposed to explain student success. Kuh et al. (2006, p. 9) note "... no single view is comprehensive enough to account for the complicated set of factors that interact to influence student and institutional performance ..." Student performance and grades are generally explained by the school climate including instructor characteristics and student preparedness and motivations (Kuh et al. 2006; Wang and Degol 2016). Two good reviews are Wang and Degol (2016) at the pre-university level which provides a comprehensive conceptualization including many factors also relevant at the university level and Kuh et al. (2006) which reviews university level studies with emphasis on student preparedness and behaviors.

Based on previous research, factors influencing grades are categorized into three groups, institutional, instructor-specific, and student characteristics. Institutional factors under university, college or departmental control include educational polices and curricula, resources devoted to teaching, and structural features ranging from quality of the classroom to class features. Instructor's factors include the experience and ability of instructors to convey information to students along with incentives associated with their position. Kuh et al. (2006) note theories of student success in college emphasize the importance of academic preparedness and experiences during college. Variables for these student characteristics, however, are difficult to measure or obtain especially over time. Proxies are commonly used. Further, to help account for institutional and instructor specific characteristics, three-stage hierarchical models with two entity-specific levels are estimated (one model is developed for class mean GPA and the other for class GPA standard deviation). The two levels are department and instructor. Previous studies have also used fixed effects models to control for unobservable effects (Kokkelenberg, Dillon, and Christy, 2008; Beenstock and Feldman, 2016; Hernández-Julián and Looney, 2016). No study was found that explicitly considered only the college of agriculture with the level of detail in the current study.

Grade inflation and distributional changes

Juola (1979) shows grade increases of approximately 0.4 grade points using data from 1965 – 1973 with the largest increases occurring between 1968 and 1970. Studies using data from the 1970's and early 80's show either a decrease or only a small increase in GPA (Juola, 1979; Bejar and Blew, 1981; Adelman, 2004). Later studies suggest grade inflation renewed in the late 1980s (Kuh and Hu, 1999; Rojstaczer and Healy, 2010; Kostal, Kuncel, and Sackett, 2016). Grade inflation leads to changes in grade distributions. Suslow (1976) notes there has been a shift in grade distributions with a decrease in the proportion of C's, D's, and F's and an increase the proportion of A's and B's. Rojstaczer and Healy (2012) suggest that since the early 1980's the largest change in grade distribution is an increase in proportion of As with corresponding decrease in Bs and Cs. The proportion of Cs shows the largest decrease with the proportion of Ds and Fs also showing a slight decrease. The decrease in Ds and Fs may be a function of increasing withdrawals from classes. Adelman (2004), for example, mentions that an important issue in change in the distribution of grades is a doubling of the percent of withdrawal or noncredit between 1972 and 1992 cohorts. Jewell, McPherson, and Tieslau (2013) suggest grade inflation may be nonlinear; they find trend and trend squared variables are significant in explaining grades. In contrast, Kezim, Pariseau, and Quinn (2005) suggest a linear trend in grades over a 20-year period for a private college business school.

Not all studies, however, agree with these findings of grade inflation, suggesting grade inflation does not exist or is a non-issue (Adelman, 2008; Brighouse, 2008; Pattison, Grodsky, and Muller, 2013). McAllister, Jiang, and Aghazadeh (2008), for example, show increasing grades in an engineering department match increasing achievement potential as measured by standardized test scores. Along these lines, Kostal, Kuncel, and Sackett (2016) note it is hard to indicate the extent of grade inflation, because of the many other factors that can contribute to raising grades; one must remove the influence of these factors to determine the severity of grade inflation. Kuh and Hu (1999) explore grade increases after controlling for student characteristics, such as family educational background, academic achievements in high school, and socioeconomic status. They show grade increases net of these specific controls is slightly larger than without controls. Kostal, Kuncel, and Sackett (2016) results, however, show an opposite effect of including control factors. These studies suggest the need to control for institutional, instructor, and student factors when examining potential grade inflation.

Institutional factors

Studies such as Diette and Raghav (2015) and Hernández-Julián and Looney (2016) use fixed effects models to show there are differences associated with students' grades and their major. Danilowicz-Gösele et al. (2017) suggest one reason for differences in departments maybe student self-selection. Further, Diette and Raghav (2015) and Hernández-Julián and Looney (2016) discuss different departments have different grading norms that may be used to manage demand for courses and majors. Bond and Mumford (2019) examine increase in grades across various colleges at Purdue University. Their results suggest differences exist between colleges. They show much of the increases in grades in the college of agriculture is driven by increasing student ability. Department / colleges characteristics and differences, however, are difficult to quantify. As such, to account for department specific characteristics, departments are added as a level in the models.

Empirical studies and theoretical reviews find quantifiable variables under institutional control are important in student success (McElroy and Mosteller, 2006; Cuseo, 2007; Kokkelenberg, Dillon, and Christy, 2008; Henebry, 2010; Wang and Degol, 2016). Cuseo (2007, p. 12) in his review on class size concludes

> The research reviewed in this article indicates that large class size is a contextual variable that has generally adverse effects on student learning, mediated primarily by lowering students' level of engagement (active involvement) with the course instructor, with classmates, and with the subject matter.

He goes on to state the only argument for large class size is fiscal. Class meeting times may also influence learning. McElroy and Mosteller (2006) believe "optimal learning times" depends on the students' circadian rhythm. Skinner (1985) finds class mean GPAs are lower for morning classes compared to afternoon and evening classes. Henebry (2010) finds students are more apt to pass a class that meets more than once a week. Skinner (1985), however, finds an opposite effect on the number of class meeting times. He suggests better attendance in classes meeting only once a week may improve grades. Student performance differs for length of class and meeting times be it because of student circadian rhythms or instructor ability.

Instructor-specific factors

The importance of the instructor and instructional method in students' learning is well documented (Kezim, Pariseau, and Quinn, 2005; Kuh et al., 2006; Joyce et al. 2014). Hoffmann and Oreopoulos (2009 p. 491) note “Hard-to-measure instructor qualities may matter more in predicting achievement, even for instructors that exhibit the same age, salary, rank, and gender.” Given the importance and difficulty of measuring these qualities especially for data going back to 1985, instructors are modeled as a level in the mixed effect models.

In addition to individual instructor effects, other quantifiable characteristics are shown in the literature to influence grades. Studies (Moore and Trahan, 1998; Ronco and Cahill, 2004; Kezim, Pariseau, and Quinn, 2005; Sonner, 2000) suggest grade inflation may be related to the increase in the use of non-tenure track faculty (term used loosely here as different studies used different terminology and length of contract). Kezim, Pariseau, and Quinn (2005) and Sonner (2000) indicate the use of non-tenured faculty is a cause of grade inflation, whereas Figlio, Schapiro, and Soter (2015) find students learn more from non-tenured faculty, which would not be grade inflation. On the other hand, Chen, Hansen, and Lowe (2021) state instructors hired on a temporary, part-time basis assign higher grades than full-time instructors without a discernable increase in learning. Hoffmann and Oreopoulos (2009) and Solanki and Xu (2018) find instructor gender plays a small but statistically significant role in student achievement. Both Hoffman and Oreopoulos (2009) and Solanki and Xu (2018) indicate a same sex instructor increases grade performance. Kapitanoff and Pandey (2017) suggest at least for the first exam, females underperform at a higher rate than males if the instructor is also female. Stratton, Myers and King (1994) find as instructors’ experience measured in years of teaching increases, grades also increase. In summary, both quantifiable and unquantifiable instructor characteristics have been shown to influence student’s performance.

Student-specific factors

Kuh et al. (2006) review theoretical perspectives on student learning including sociological, organizational, psychological, cultural, and economic perspectives. They conclude the students "... wide path with twists, turns, detours, roundabouts, and occasional dead ends ..." is more realistic than a direct route to student achievement (Kuh et al., 2006 p. 7). In their framework, besides institutional factors, pre-college experiences and student behaviors or engagement also matter in student success. Pre-college experiences include enrollment choices, academic preparations, aptitude, college readiness, educational aspirations, family and peer support, motivation to learn, gender, race, ethnicity, and demographics. Of these experiences, Kuh et. al. (2006, p. 31) conclude, "In fact, the best predictor of college grades is the combination of an individual student's academic preparation, high school grades, aspirations, and motivation." Kobrin et al. (2008) and Westrick et al. (2015) indicate the importance of pre-college preparation in explaining the first-year success of college students. Rothstein (2004), however, finds much of the SAT predictive power on grades is derived from SATs correlation with high school demographics. Studies also suggest changing demographics may lead to higher grades (Kuh and Hu, 1999; Kostal, Kuncel, and Sackett, 2016). Kuh and Hu (1999) suggest increased female enrollments in colleges may lead to higher grades because females tend to receive higher grades, while Bergtold, Yeager, and Griffin (2016) did not find any significant differences between performances of male and female students.

Although brief, this review shows the importance of controlling for factors that may influence grades; therefore, creating a better picture of whether grade inflation is occurring. Further, it is not just the institution, instructor, or student factors that determine student success, but rather the interaction of all three of these factors.

Data and empirical model

Data

Data for 13 departments and one study program in COALS from Spring 1985 through Fall 2019 (the last pre-COVID 19 semester) are utilized giving 35 years of data, far more than many studies. The data starts in Fall 1985 because this is the first semester grade data are available. The primary reason behind selecting COALS is access to the data. Another important factor is the wide range of diverse disciplines within COALS which range from basic science and technology to economics and agricultural communication. This wide spectrum of disciplines helps make the findings relevant to other universities. In addition, methods utilized are applicable to other universities. Data from the University are supplemented with information from undergraduate catalogs, faculty and department websites, e-mails to instructors, as well as discussions with staff and faculty in the various departments. Departments changing names and combining departments resulted in no data for Forestry and Rangeland Science Departments. A total of 17,696 usable observations are obtained. All data are at the class level.

Grading scheme used by TAMU is a four-point scale based on receiving four points for an A, three for a B, two for a C, one for a D, and zero for an F (TAMU does not use grade modifiers of plus or minus). Class means GPA are based on the number of students receiving each grade. Grade distribution is based on the standard deviation of the grades received in the class. Individual projects, study abroad, and internship classes are not included. These classes are eliminated because students tend to receive A's or credit for completion in these classes. Distant learning or in-absentia classes are eliminated because they do not have a specific meeting time and number of classes per week, variables used in the model. Classes with variable credits are also eliminated. Further, for privacy reasons, information for any classes of less than five students is not available. No exact information on the number of classes that fall into this category is available. Summer classes are also eliminated from the analysis, because most summer courses, especially early in the data set, fall into one of the categories of classes that are dropped. Further it is felt, summer courses being shorter in length, either five or 10 weeks, instead of 14 or 15 weeks (changed over the data set) may not be directly comparable. Data from Texas A&M were given by sections. In some cases, classes in COALS might have different sections taught at the same time by the same instructor. Sections that were held at the same time and by the same professor were merged to reflect the fact that the sections meet as one class. The different sections are generally for accounting and reporting purposes and there is not a difference in the class delivery.

Student grade information is highly protected by students' privacy rights laws. Because of this privacy, this research uses class level data rather than individual student data. Over the data period, changes occurred in the data reporting formats and requirements from year to year, specifically in student admissions information and information reported in university catalogs. As such, in many cases, department and administration leadership were contacted to understand data changes. Further, much of the faculty information was obtained by internet searches or contacting instructors directly. These sources of information were particularly used for early years and nontenure track instructors.

Variable descriptions and summary statistics

Based on the literature, a theoretical model of the three sets of factors influencing grades is presented in Figure 1. Institutional, instructor, and student factors come together to determine the class mean GPA and standard deviation.

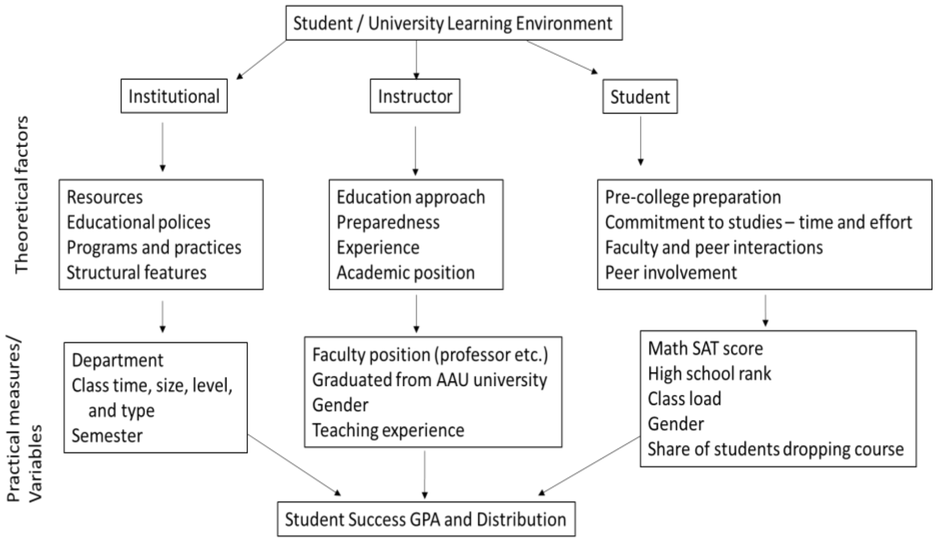


Figure 1. Institutional, Instructor, and Student Variables that my Affect a Students' Grade

Empirical variables used as proxies for the theoretical model, their descriptions, and summary statistics are presented in Table A1 in the Appendix. Most of the variables are self-explanatory. Class averages for variables that are based on student information, such as SAT scores are based on all students in the class that had such information reported. TAMU, COALS, and departments early in the data set did not require students to provide SAT scores or their rank in high school for admission to the University. As such, the class average may be based on less students than in the class, especially in the early years. To account for this non-reporting, a qualitative 0-1 variable for the years 1985-1988 is included with one indicating the years 1985-1988. This variable is then interacted with SAT and High School rank. Variables associated with Association of American Universities (AAU) and foreign universities are included as a proxy for potential differences in instructor background (not quality). Association of American Universities is comprised of some of the leading universities in the U.S. and Canada.

Qualitative variables' (listed as 0-1 variables in Table A1) mean values can be interpreted as the share of class in the category. As to institutional variables, most classes meet either in the morning or afternoon, meet twice a week, and are upper division classes. Average class size is 51 students but ranges from five to 349 students. Approximately 78 percent of the instructors are male. Few instructors' last degree is from a foreign institution. At the time of their graduation, slightly more instructors received their degree from non-AAU affiliated institutions than from AAU affiliated universities. Average percent of male students in class is slightly higher than females. SAT scores are somewhat above the national average (College Board, 2015). High school rank average is approximately the 79 percentile (or top 21% of the graduating class in the high school). Students took on average 14 credits per semester with only a small percentage of students dropping or receiving a no grade in a class.

In Figure 2, mean GPA and standard deviation are graphed by semester for all classes. Mean weighted GPA overall years is 3.08. In Fall 1985, mean weighted GPAs is 2.88. By Fall 2019 GPA has increased to 3.28. Standard deviation declined from 0.85 to 0.73.

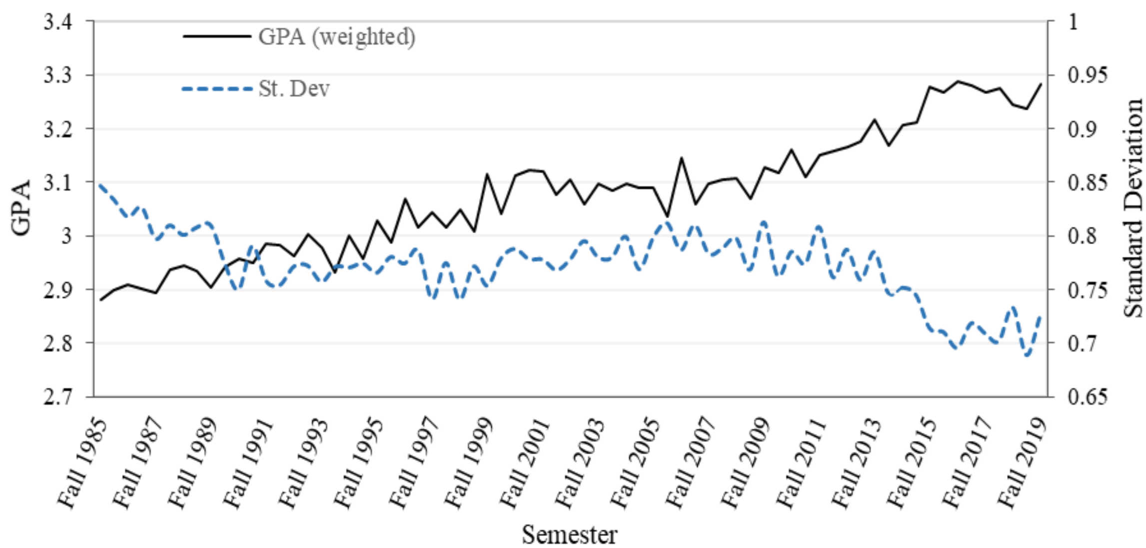


Figure 2. Weighted (by class size) Average COALS GPA and Standard Deviation by Semester from Fall 1985 to Fall 2019

Examining the percentage of letter grades helps explain the above increase in GPA and decrease in standard deviation (Figure 3). The percentage of A's increases from about 32% to over 52%. All other letter grades percentages have decreased over the time frame with the percentage of C's showing the largest decrease (from approximately 22% to a little more than 13%).

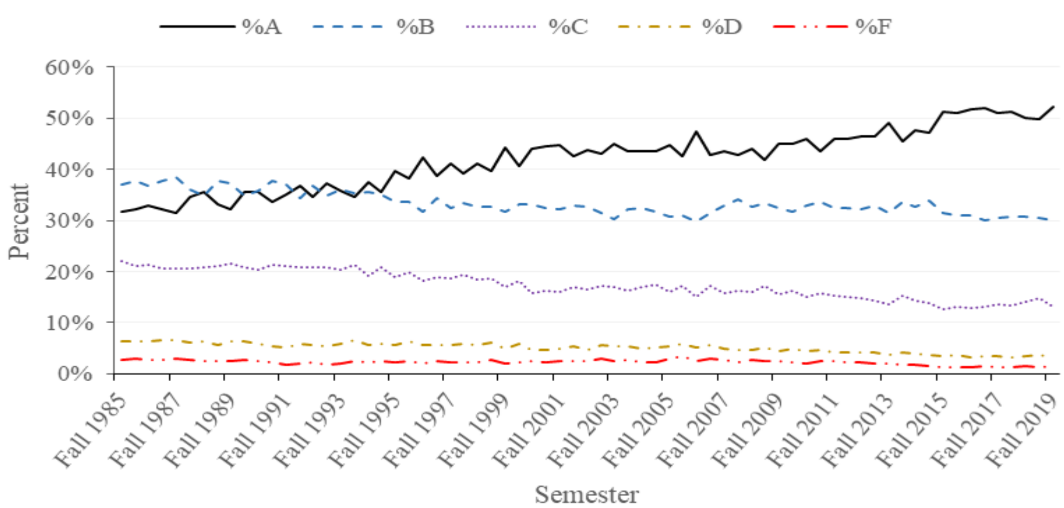


Figure 3. Percentage of Letter Grades in COALS in all Courses from Fall 1985 to Fall 2019

Mean GPAs and standard deviations by department show considerable variation, with mean GPAs, ranging from 2.79 for the Renewable and Natural Resources study program to 3.46 for the Department of Agricultural Leadership Education and Communication (Figure 4).

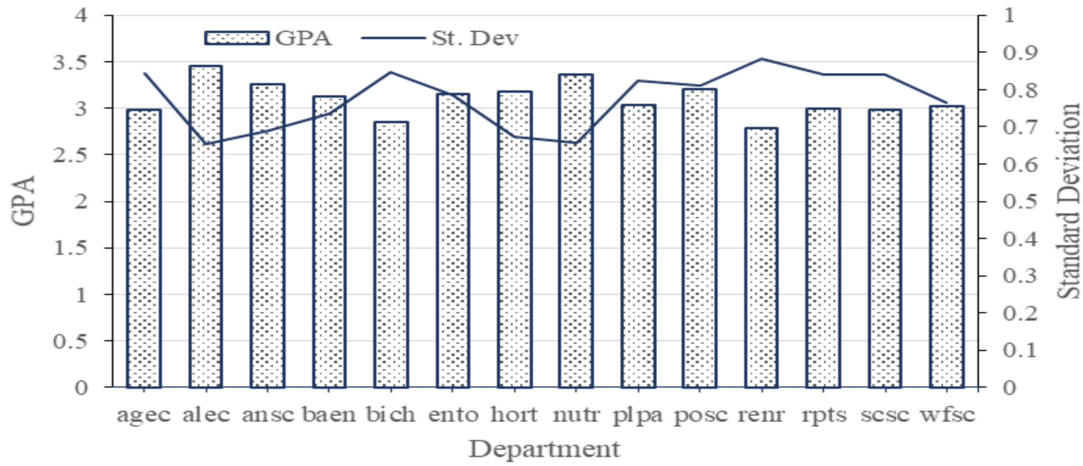


Figure 4. Overall Weighted by Student Mean GPA and Standard Deviation by Department in COALS for the Fall 1985 to Fall 2019

Standard deviation and GPA show an inverse relationship caused by the increase in A's and decrease in F's. Such departmental differences provide support for the inclusion of departments as a level in the model.

Estimation methodology

When discussing observations that are a part of one group or a cluster there is risk of drawing inferences about the groups based on simple aggregation of the individual data. Diez-Roux (1998) argues the relationships observed at the individual level may not hold between the same variables observed in a group, which will result in inferential errors. At the group level, there may be specific group properties that will affect the variables in the same way, regardless of the individual characteristics. In the scope of this study, there may be department specific characteristics that might follow the same pattern within the same department. This means that data within the same department might be correlated, whereas data from different departments may be considered independent. A similar argument holds for instructors.

Multilevel mixed modeling accounts for both within-individual and between-individual variations (Douglas, 2004; Wu, 2019; Shinn and Rapkin, 2000). Following Kokkelenberg, Dillon, and Christy (2008), Beenstock and Feldman (2016), and Hernández-Julián and Looney (2016), a mixed effect model is used to incorporate the variations of grading between the departments and individual instructors, as well as to account for the internal correlation while controlling for other factors. The model is a three-stage hierarchical model with two levels, where the first stage measures the fixed effect or within-individual variation, while the second and third stages measure the random effect or the between individual variations for instructors and for departments. The estimation consists of separate sets of regressions nested within each other (Greene, 2012). Combining the stages, one obtains

$$y_{ijk} = \beta_{0jk} + \sum_{m=1}^{k} \beta_{mjk} X_{mjk} + \varepsilon_{ijk} \qquad (1)$$

y_{ijk} - is the vector of observations (grades) for the i^{th} class taught by the j^{th} instructor, nested within the k^{th} department

X_{mjk} – is i^{th} class characteristics (see variables in Table A1),

β_{0jk} - is the vector of fixed effects, or mean GPA for the j^{th} instructor within k^{th} department

β_{mjk} – is the coefficients for i^{th} class individual characteristics,

ε_{ijk} - is the vector of residuals for the i^{th} class taught by the j^{th} instructor within k^{th} department

Level 2 equation is

$$\beta_{0jk} = \gamma_{00k} + u_{0jk} \qquad (2)$$

where

β_{0jk} - is the vector of fixed effects or mean GPA for the j^{th} instructor within k^{th} department

γ_{00k} - is intercept for the k^{th} department

u_{0jk} - is deviation of the j-th instructor from the intercept of the k^{th} department

Level 3 equation is

$$\gamma_{00k} = \theta_{000} + \varepsilon_{00k} \tag{3}$$

where

γ_{00k} - is intercept for the k^{th} department

θ_{000} - is fixed intercept across all groups

ε_{00k}- is deviation of the k^{th} department from the fixed intercept.

The model is estimated using mixed command in Stata 15 (StataCorp, 2017).

Results and discussions

Estimation results for both the GPA and standard deviation (StDev) models are presented in Table A2 in the Appendix. Wald statistics for both models indicate jointly all variables except the intercept differ from zero at p-values less than 0.01. Likelihood ratio tests reject the null hypotheses the models including random effects are equivalent to a linear regression model. Evidence of clusterization of the data is tested using interclass correlation coefficients, which are the ratio of one variance to the sum of the other two variances. The coefficient for level two, for example, is the ratio of level two estimated variance to the sum of the level one and level two estimated variances. Because this measure is greater than 0 (0.14 for the department and 0.55 for the instructor given department), the conclusion clustering is indicated (Park and Lake 2005). The selection of both levels (random effects) is also supported by model selection information criteria, minimum values were attained using both levels when compared to one level or no level.

Finally, both random effects parameters differ from zero in the two models. As is common practice, the coefficients associated with the levels are not presented given the number of coefficients, rather just over all significance is discussed. Overall, model statistics provide evidence for the appropriateness of the mixed effects model.

Trend variables

In the GPA model, the trend variable is significant and positive, indicating over the period 1985 to 2019, there is evidence of systematic grade increases in COALS (Table A2). Without the department and instructor levels included in the model, trend remains positive, significant, and is larger in magnitude. Department and instructor levels appear to partially account for the increases in GPA seen in Figure 2. In the StDev model, the trend variable is significant and positive. Without the department and instructor levels, the coefficient on trend in StDev model is not significant.

Institutional variables

Estimated institutional random-effects parameters are significant in both the GPA and StDev models indicating common grading patterns within a department. These results are in line with previous studies such as Diette and Raghav (2015), Hernández-Julián and Looney (2016), and O'Connor (2020). Variables associated with Y85-88 indicate university reporting requirements may help explain GPA. All other institutional variables in the GPA model except semester, upper division classes, and meeting twice a week are significant at p-values of 0.01 or less indicating institutional variables affect student learning as measured by GPA. Semester being insignificant indicates class GPA does not vary by Spring and Fall semesters, a pleasing result for COALS. GPAs for classes held twice a week do not significantly differ from classes held three times a week. However, GPAs for classes meeting once a week are higher than those meeting three times a week. Classes that meet more often per week are generally shorter in duration. Contrary to Henebury (2010), longer classes appear to result in higher GPAs. This result may be because longer duration classes tend to more specialized attracting interested and motivated students.

Classes held during the morning hours result in lower GPA than classes that are held in the afternoon and in the evening. Based on a Wald test, grades for evening and afternoon classes do not significantly differ from each other. One possible explanation for morning classes having lower GPAs is students in morning classes may not have had enough sleep, which affects their performance during the class (Wahlstrom, Hendrix, and Frederickson 1998; Wheaton, Chapman, and Croft 2016). Further, evening classes tend to be more specialized courses that may be contributing to larger GPAs. There is no significant difference between grades in upper and lower division classes. Classes with three or more credit hours tend to have lower GPAs than classes with one or two credit hours.

GPA is negatively correlated with class size, which is consistent with previous findings (Kokkelenberg, Dillon, and Christy, 2008). Higher grades for smaller classes may be a result of better learning outcomes (Nye, Hedges, and Konstantopoulos, 2000), because in smaller classes instructors are able to provide more time to individual students both during class and office hours (Jewell, McPherson, and Tieslau, 2013). All institutional variables' coefficients in the StDev model are significant at p-values of 0.01 or less except those for semester and classes that meet twice a week. The signs of the significant coefficients in the StDev model tend to be opposite of the signs in the GPA model. Generally, variables that increase GPA tend to decrease standard deviation or variability of grades. Most likely cause of this inverse relationship is GPAs having an upper bound of four.

Instructor-specific variables

As in the case with departments, the random part of the model indicates significant variability in grading patterns of individual instructors. Coefficients associated with associate professor, lecturer graduate student, and lecturer other are significant and positive indicating classes taught by these instructors on average have higher GPAs than those taught by a professor. The increasing trend of using graduate students and other lecturers accounts for some of the increase seen in GPAs overtime. The questions are do students learn more from these non-professorial rank instructors, which would not be grade inflation, or do these instructors grade easier, which is grade inflation? Unfortunately, the data set cannot answer these questions. Anecdotal evidence provided by some graduate students suggests their incentive is to get good teaching evaluations to help in obtaining the first faculty position. Easy grading resulting in higher GPAs may contribute to higher evaluations (Pease, 2017, Howard and Maxwell, 1980; Eiszler, 2002). Kezim, Pariseau, and Quinn (2005) and Sonner (2000) also find visiting and adjunct faculty grades are higher. They suggest that the reason for the higher grades may be the expectation of higher student evaluations because these instructors are often hired on the term-by-term basis and higher student evaluations are more likely to result in their contract being extended. Classes with instructors who graduated from an AAU university have higher GPAs than classes with non- AAU or foreign universities' instructors. Instructor gender does not significantly influence GPAs. This result differs from some previous studies (Moore and Trahan, 1998; Kezim, Pariseau, and Quinn, 2005) and is in line with others (Figlio, Schapiro, and Soter, 2015; Ronco and Cahill, 2004). The insignificance may arise from controlling for individual instructor variability.

In the StDev model compared to professors, assistant professors tend to have more variable GPAs, while other faculty members show smaller variability in grades. Assistant professors having a larger variability may be because of differences in experience and the number of times an instructor has taught a particular class.

Student characteristics variables

All student variables are significant. Coefficients associated with variables measuring student motivation and ability, SAT, class load, and high school rank, are positive. These results are supported by previous studies, such as Kobrin et al. (2008) and Westrick et al. (2015) who find high school performance and innate ability of students are good indicators of university performance. Student motivation to achieve and less time to procrastinate may explain why larger class loads results in slightly higher GPAs. As the percent of no grades increases, the class GPA decreases, which may be because student tend to drop harder classes or classes with more work to maintain their GPA. As the percent of females in the class increases (percent male decreases) GPAs increase. Kuh and Hu (1999), and Voyer and Voyer (2014) find females tend to have higher grades than male students.

Both the number of female students in COALS and the students average SAT have been increasing over time (Figure 5).

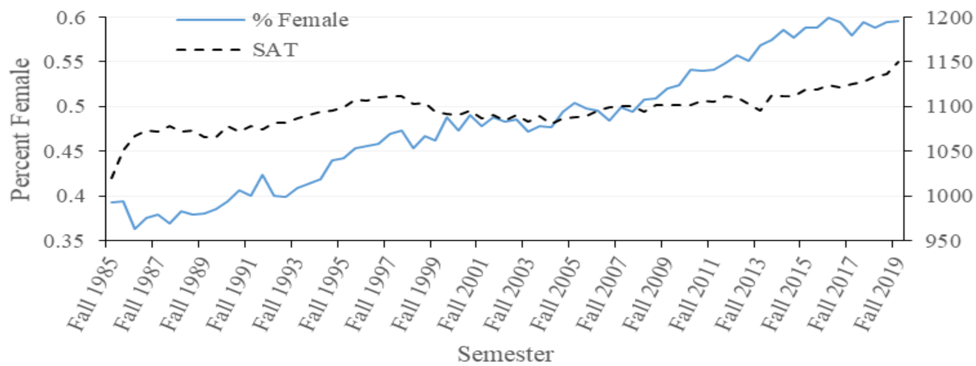


Figure 5. Average SAT Score for all Students in COALS and Percentage of Female Students by Year from Fall 1985 to Fall 2019

These increases help explain the increasing grades in COALS. Increasing SAT would not cause grade inflation, because better students are expected to obtain better grades. The reason females tend to have better grades determines if the increased percent of females is associated with grade inflation. If instructors simply grant higher grades to females, then grade inflation is present; however, if females are more motivated and study harder, it is not a grade inflation. Unfortunately, the data set cannot clarify this issue. Dampening the GPA increase is student load and high school rank, both of which decrease slightly overtime.

All student variables are significant in the StDev model. There is higher variability in grades for classes with more male students than female students. The higher the student load, SAT score, and high school rank, the lower the variability in grades.

Excluding early years

Grade inflation is considered an empirical issue occurring over decades, as such the analysis used all years in which data were available. To examine if problems with data consistency in the years 1985-1988 impact the results, the model is re-estimated excluding these years. In removing the years, 1,735 observations are lost. All coefficients' signs are unchanged. Both the trend and lecturer graduate coefficients are less significant but still significant at 6% or less. As such there is no real change in inferences between including or excluding the early years.

Conclusions and Discussions

As discussed earlier, grade inflation is important to universities, instructors, and students. Universities competing for reputation, students and resources may permit grade inflation to occur (Jewell, McPherson, and Tieslau, 2013). Instructors may inflate grades for personal gain such as favorably biasing student evaluations (Jaschik, 2008; Pease, 2017), whereas students may pressure academics for higher grades (Peace, 2017). Grade inflation, however, not only affects universities, instructors, and students, but also spills over into the employment sector where the recruitment of a good candidate is often based on university GPA (Adams, 2015; Fossati, Wilson, and Bonoli, 2020). A high GPA may suggest a job candidate has high aptitude and motivation, but on the other hand it could just be the result of inflated grades. It is clear comparing student GPAs across the time in the presence of grade inflation is inappropriate and may result in poor hires by companies. Whether grade increases are grade inflation or caused by improvements in teaching and students' preparedness is an empirical issue. As such using data from Texas A&M College of Agriculture and Life Sciences (COALS) three interrelated objectives are to examine: 1) has grade inflation occurred, 2) has the distribution of grades as measured by the standard deviation changed; and 3) how various factors influence class mean GPA and grade distributions. Even though the study is conducted based on data from a specific institution, the results may be generalized to other land grant universities, because the results support findings of earlier studies that did not explicitly focus on colleges of agriculture.

Utilizing data from 14 departments in COALS for the years 1985-2019, this study found evidence of systematic increases in mean class GPAs and standard deviation of grades after controlling for the influence of institution, instructors, and student factors. The results are suggestive of grade inflation, but with the caveat not all factors potentially influencing grades could be included. All three groups of factors are important in explaining class GPAs and standard deviations. Overall, the findings and results are quite robust to changes in model assumptions. If the model is simplified to one level (department) or to no levels, majority of conclusions about the inferences of the various characteristics on mean class GPAs and standard deviation hold; supporting the validity of model.

Findings have implications on recruitment policies and requirements. With respect to instructors, a positive signal to administration is seen in that there are no differences in GPAs by instructor gender. The practice of increasingly hiring graduate students and other instructors is associated with increasing grades, but is it inflation or do these instructors provide a better learning environment? Other instructors, for example, tend to have less of a research role, which leaves more time to allocate to class preparation and lecturing. If inflation, the trend of hiring non-tenure track instructors may be particularly worrisome as universities move towards increasing use of non-permanent faculty. Addressing this increase in grades, however, is not easy given an environment of decreasing budgets and increasing accountability as measured by evaluations of instructors' performance. Studies have suggested grades and evaluations are positively correlated (Marsh and Roche, 2000). If that is the case, then other measures of instructor effectiveness may have to be used. Hiring instructors with AAU backgrounds may result in higher GPAs but the reason is not clear. More research is necessary here; is it, for example, better preparedness by instructors or more focus on research that results in different class GPAs. Assuming GPAs are good measures of learning, administrators may choose to schedule their core, degree-specific classes later during the day to maximize student learning. Further, smaller class sizes appear to be beneficial for learning. Universities should stride for smaller classes instead of larger classes. Budgetary considerations, however, are driving both class size and use of graduate students and other instructors. Some larger classes, however, may serve to weed-out students from the major, instructors in these classes may give lower grades (Weston et al., 2019). This observation complicates the issue.

From an instructor standpoint, implications are few. Given the trend of using student evaluations and assuming higher class GPAs are associated with better evaluations, instructors may want to ask to teach smaller class sections of specialized courses. Another consideration is class duration. Longer classes yield higher GPAs, but this is most likely because longer and less frequent classes tend to be more specialized.

With the goal of potentially higher GPA, when selecting between different sections of a class, students could check the instructor background, and should try to register for the sections that have fewer students. In addition, given a chance, they may consider taking classes offered later during the day. Students may take classes taught by graduate students and other instructors in anticipation of higher grades but should be aware of the potential opportunity cost of missing on insights when taught by more experienced professorial track faculty members.

The results suggest further research on separating the effects of grade improvement and grade inflation is necessary. Because variations are seen by department, examining departments separately or by general focus such as STEM or social is seen as a valuable contribution. Improved understanding of this phenomena and identifying source of grade increases is essential to determine if grade inflation is occurring. Research on grading patterns directly before and after instructors' promotion would provide insights on how the promotion process impacts grades. Whether instructors shift their time and effort towards publications and away from teaching or if they grade easy to improve evaluations needs exploring. Although, different colleges of agriculture have somewhat different employment areas, they do compete against each other for hiring of their graduates. Further, how graduate schools and employers use / compare students' GPAs from different university should be explored, given the empirical nature of grading. Can or should grades from different universities be compared? Being an empirical issue, how do grades in colleges of agriculture vary by university. Such research would benefit both graduate colleges and employers. How has COVID-19 change the grading environment? If changes occurred, are they permanent changes?

As with all studies, limitations exist. Being a case study, all limitations of such studies apply here. The primary limitation is the inability to include all characteristics influencing class GPAs. Instructor level may not account for all influence of an instructor, such as changing teaching styles. Unfortunately, such information is not available going back to 1985. An attempt was made to include infrastructure (building, classroom, lab modernization, electronic equipment available, etc.) in each classroom and how it evolved over time. Such details, however, are not available for all the classrooms. Including interaction terms may be appropriate, but with the number of variables in this study the question becomes which interactions, because including all interactions creates an unwieldly model. Including only some interactions may create a data mining exercise.

ARE WE THAT SIMILAR? DIFFERENCES IN GRADING PATTERNS WITHIN THE SAME COLLEGE

Introduction

College graduates have an advantage over nondegree holders in many sectors of the economy (Schaeffer, 2022); despite this advantage only approximately 39% of the work force holds an associate or bachelor's degree (US Census Bureau, 2022). College graduates have lower unemployment rates and larger median annual salaries. The gap between annual salaries is growing, increasing by almost $9,000 between 1990 and 2021 (Schaeffer, 2022). College enrollment peaked in 2010, with enrollment in 2021 being over 2 million less students than in 2010 (Hanson, 2022). Other issues are only about 60% of students who enrolled in a college earn a degree within six years, and differences in enrollment and graduation rates differ by income level, race, and first-generation student (Causey et al., 2022).

In response to the above concerns, the Biden administration proposed the College Completion Fund Act of 2021 which would establish "… a grant program for participating public institutions of higher education to provide student support services to increase participation, retention, and completion rates of students from low-income backgrounds, historically underrepresented students, first-generation college enrollees, parenting students, students with disabilities, and student veterans" (GovInfo, 2021). The bill's intent is to increase college access and affordability and ensure more students complete college and enjoy the benefits of a college degree. The importance of completion rates is stressed in the bill. Denning et al. (2022) provide reasons why one might see decreasing graduation rates such as increasing tuition costs, increase in hours worked by students, less time spent studying, and decreasing standardized math scores (which may be evidence of students' being less prepared for college). They, however, note completion rates have been increasing rather than decreasing partially because of increasing grade point averages (GPAs). Although grade inflation (an increase in student grades without an associated increase in knowledge and learning) may be addressing the social problem of low completion rates but at the costs of potential declining college wage premium associated with decreased learning (Denning et al., 2022).

Another issue is passage of such legislation may create an enabling environment for recruitment and retention of students in universities. Competition for students and increasing tuition and costs may lead universities to allow grade inflation in the hopes of increasing their reputation, increasing enrollment, and justifying tuition increases (Jewell, McPherson, and Tieslau, 2013). Attracting additional students may also provide the university more funds through tuition and fees (Teixeira et al., 2014), but higher than deserved grades will eventually negatively affect the reputation of a university for failing to prepare professionals that meet industries' expectations (Chowdhury, 2018).

Hermanowicz and Woodring (2019, p. 497) note "Grades are a ubiquitous part of college" influencing a large part of undergraduate life from self-definition to graduation and job prospects (Becker, Geer, and Hughes, 1968; Rojstaczer and Healy, 2012). With grades being such an important part of undergraduate life, it is no surprise studies such as Butcher, McEwan, and Weerapana (2014), Minaya (2020), and Opstad (2020) show grades may influence students' choice of majors. Further, different grading norms can be used to manage demand for majors (Diette and Raghav, 2015; Hernández-Julián and Looney, 2016). Several studies have shown grading norms may differ between universities, colleges, and even different departments within a college in a university (Hartnett and Centra, 1977; Achen and Courant, 2009; Herron and

Markovich, 2017; Bond and Mumford, 2019). Although these studies suggest there are differences in grading patterns between departments of a college or university, they provide no clear evidence on factors causing these differences.

Using a unique dataset, hierarchical mixed effect models are used to identify factors influencing class level GPAs in departments in College of Agriculture and Life Sciences (COALS) at Texas A&M University (TAMU). Interviews with different departments' personnel and comparative analysis of exogenous factors are implemented to better understand grades overtime. Specifically, the objectives are to:

1) determine if potential grade inflation has been occurring by department and if it differs overtime, and
2) examine factors influencing mean class GPA among different departments in COALS to provide information on factors causing these differences and explore if the influence has changed overtime.

This study contributes to the existing literature in that it considers a wide array of factors affecting grades in different disciplines and draws parallels among departments. Grades over time and differences in factors affecting grades may be used by the departments' administrations to understand whether the changes in grading patterns are the result of improved learning or consequences of inflated grades.

Literature Review

In the past decades, there has been intense competition among universities for high school graduates (Voice of America, 2019). One reason for this competition is decreasing government spending on public education (Cattaneo et al., 2016) which forces universities to attract funding through additional sources including student tuition and fees (Teixeira et al., 2014). Universities must either increase tuition and fees or enroll more students to address budget shortfalls. Both tasks are challenging. Justification for tuition and fees increases includes improved services and quality of education, which often leads to additional expenditures (Archibald and Feldman, 2012). In addition, students' mobility and geographic integration of college markets (Hoxby, 2000), as well as emergence of online education make attracting additional students harder. In this competitive environment, some universities try to increase their image and reputation while others, rather than engaging in expensive competition, simply accept weaker students (Jefferson, Gowar, and Naef, 2019). Peace (2017) argues that even weaker students expect good grades in return for high tuition and fees. This notion of "consumerism" creates pressure on institutions to grant higher than deserved grades. Instructors as well may be inclined to grade leniently to avoid time consuming arguments with students, especially on the assignments that may not have a right or wrong answers (Achen and Courant, 2009). There is also a labor market justification for granting higher than deserved grades. Graduates from disciplines with higher paying jobs generally have lower grades compared to those graduating from lower pay job disciplines. This grade disparity may be used to attract students to the lower wage disciplines (Sabot and Wakerman-Linn, 1991; Freeman, 1999, Diette and Raghav, 2016).

Evidence of grade inflation and factors affecting grades are the subject of numerous studies over the past decades (e.g., Schutz et al., 2015; Kostal, Kuncel, and Sackett, 2016; Peace, 2017). Kuh et al. (2006) and Rojstaczer and Healy (2012) find that even after accounting for student aptitude, grades still increased in recent decades. Results from the first essay of this dissertation show that even after controlling for student, instructor-specific, and institutional factors, grades have increased significantly between 1985 and 2019. Denning et al. (2022) take a different approach and compare end of year test grades and students' GPAs. Over the span of 12 years, students earned better course grades in later years, although end of course exam scores stayed nearly the same (nine out of 12 exams were identical).

Grade inflation is recognized to be one of the most important issues facing the academic world (Merrow, 2004) for at least two reasons. First, inflated grades do not convey the proper message concerning students' abilities and knowledge to future employers. A student with a "B" from an institution where grade inflation is not occurring may be better prepared for the job market compared to a student with an "A" from an institution where grade inflation is occurring. Employers without knowledge of grade inflation may be tempted to hire the graduate with higher grades. Second, because grades have a cap (usually 4.0) grade inflation places a good student close to exceptional student, thus negating the ability of grades to differentiate between students even in the same institution (Kohn, 2002).

Differences in grading standards are observed not only between different universities (Popov and Bernhardt, 2013), but also between departments within the same university (Sabot and Wakerman-Linn, 1991; Herron and Markovich, 2017), and even between instructors in the same department (Jewell and McPherson, 2012). Hartnett and Centra (1977) discuss departmental differences from the standpoint of students' aptitudes and preparedness. They find significant department-specific differences in student learning outcomes. Several other studies highlight grade differences between Science, Technology, Engineering, and Mathematics (STEM) majors and non-STEM majors within the same universities (Ost, 2010; Witteveen and Attewell, 2020; Tomkin and West, 2022). One common finding is STEM departments tend to grade tougher than departments granting non-STEM degrees. Tougher grading may result in smaller number of students enrolling in STEM related disciplines (Rask, 2010). Bar, Kadiyali, and Zussman (2009) concur with this finding adding publicly available grade distributions make it possible for students to self-select into leniently graded classes. Opstad (2020) suggests students may self-select career pathways based on grades; below average performing students may select majors other low performing students select. Reason given is it may be easier for a student to obtain a good grade when competing against peer students who are also low performing or less qualified. Studies such as Hartnett and Centra (1977), Achen and Courant (2009), and Herron and Markovich (2017) highlight differences in grading patterns between departments in a college or university. Specific reasons behind the differences, as well as any suggested course of actions are usually not discussed.

Data Description and Summary Statistics

Differences in grading patterns for the following 12 departments within COALS at TAMU are examined:

- Department of Agricultural Economics (AGEC),
- Department of Agricultural Leadership, Education, and Communications (ALEC),
- Department of Recreation, Park, and Tourism Sciences (RPTS),
- Department of Animal Science (ANSC),
- Department of Poultry Science (POSC),
- Department of Biological and Agricultural Engineering (BAEN),
- Department of Biochemistry & Biophysics (BICH),
- Department of Entomology (ENTO),
- Department of Horticultural Sciences (HORT),
- Department of Plant Pathology and Microbiology (PLPA),
- Department of Soil & Crop Sciences (SCSC), and
- Wildlife and Fisheries Management (WFSC).

Data from the first essay are used, however the first essay includes only analysis of COALS as a whole and not individual departments or by different time periods. In addition, data before 1988 is excluded because of differences in students' admission requirements before and after 1988.

Descriptions of the variables used in the analysis are in Table A3 in the Appendix. Differences in variables used between the second and first essays are because of an insufficient number of observations after disaggregating the data into departments. Specifically, evening classes are merged with the afternoon classes and foreign instructors are merged with the non-AAU instructors' group.

Because what happened in the distant past may not be as relevant as the present, the data are divided into two periods, 1989-2003 and 2004-2019, which divides the data in approximately two equal periods. While differences between periods help in long term trends, the recent period may be more relevant for addressing policy changes. Thus, the comparative analysis is implemented between the departments and within each department between the two periods. GPAs are analyzed as a function of institutional (class time and duration, number of credits, upper or lower division courses, and number of total students in the class), instructor-related (instructor gender, position, and graduating from a AAU accredited university), as well as student-related (student gender, high school rank, SAT score, class load, and course completion) characteristics. Summary statistics along with tests of differences in the mean values of the variables by period are in Table A4. Finally, although written as department, it needs to be noted because the data are for class and not department, the class data may include students from multiple departments.

Departmental GPAs

Department mean GPAs show variability by department, years, and between the two periods (Figure 6). For presentation purposes, the departments are grouped into four subgroups. This grouping consists of Social Sciences, Animal Oriented, Plant Oriented, and Other.

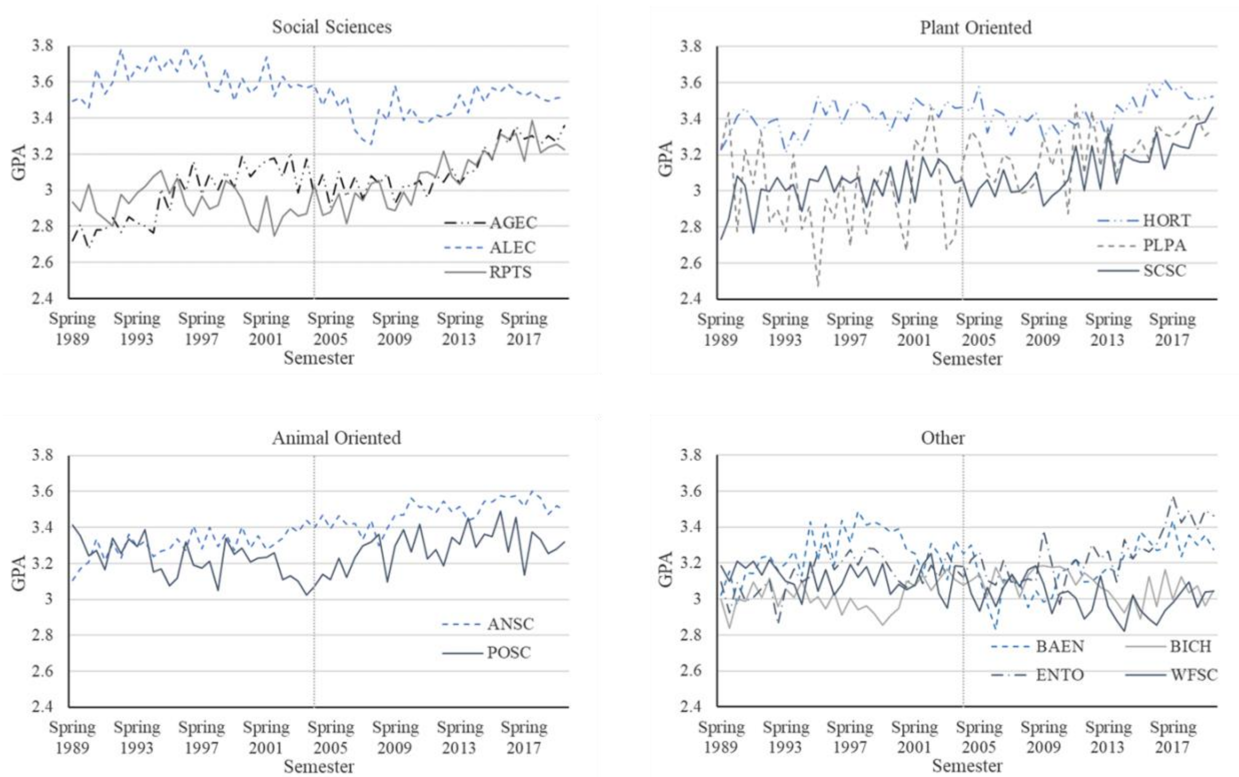


Figure 6. College of Agricultural and Life Sciences, Departments' Average GPAs by Semester from 1989 to 2019

All departments' (except HORT, SCSC, POSC) mean differences between the two period's GPAs are statistically significant (henceforth the wording significant is used for ease of reading) at p-values of 0.05 or less (Table A4). Three departments, ALEC, BAEN, and WFSC, had significant decreases in mean GPAs in the second period relative to the first period. The remaining six departments had positive significant increases. Mean departmental GPAs range from 2.91 (RPTS) to 3.61 (ALEC) in the first period, and 3.00 (WFSC) to 3.48 (ANSC) in the second period. Even within the same grouping, departments have different grading patterns. For example, within Social Sciences, RPTS's GPAs are relatively flat in the first period but show a steady increase in the second period, while AGEC's GPAs show a slight increasing trend through most of the first period then a flat or decreasing trend for the first part of the second period and an increasing trend after that until the end.

Institutional characteristics

Most classes meet in the morning. Only PLPA has less than 50% of their classes in the morning. The percentage of classes in the morning ranges from nearly 82% for ANSC and ENTO in the first period to 38% in PLPA in the second period. Except for PLPA and POSC, all departments showed a significant decrease in morning classes with the corresponding increase in afternoon classes between the two periods. There appears to be no common tendency concerning the number of times per week courses meet. Most classes in COALS are upper division (junior and senior) classes with all departments having 53% or more of their classes being upper division. In the first period, HORT and ENTO are the only departments that had nearly equal split between lower division (freshmen and sophomore) and upper division classes. However, in the second period, they increased the share of upper division classes.

ALEC, BICH, and WFSC show significant decreases in average student enrollment in classes, whereas HORT, PLPA, SCSC, ANSC, and POSC had no changes in average student enrollment per class between the two periods. AGEC, RPTS, BAEN, and ENTO had significant increases in average enrollment per class. Average class size is the largest in AGEC (over 69 students in period two) and is the smallest in BAEN (31 students in the first period). By far, most classes in COALS are three or more credits. AGEC, PLPA, and ENTO have seen decreases in the percentage of three or more credit classes. in ALEC, RPTS, SCSC, and BAEN, percentages of classes with three or more credits increased in between periods two and one.

The average number of students in class are graphed in Figure 7. There are large variations in class sizes among the departments and there is variability within a class by semester. ANSC and WFSC have stable number of students in classes, while others, such as SCSC had increases in class size till mid-2000s, then show a decrease in numbers.

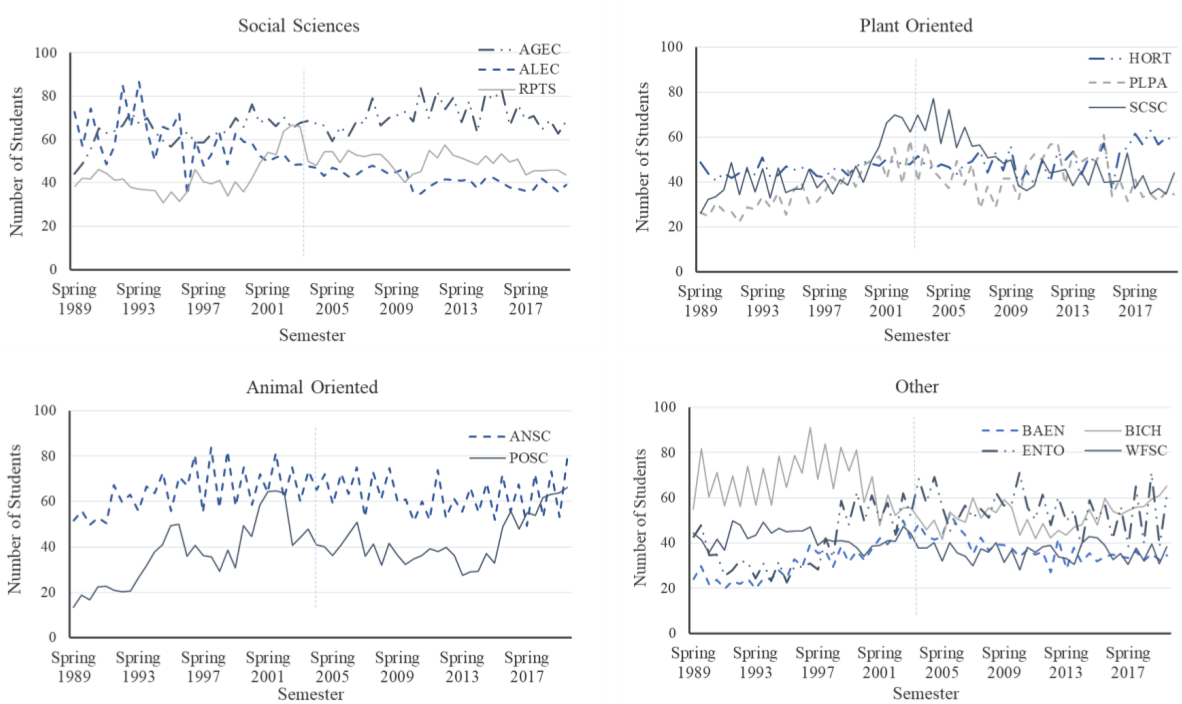


Figure 7. College of Agricultural and Life Sciences, Departments' Average Number of Students in Each Class by Semester from 1989 to 2019

Instructor characteristics

All departments have significant decreases in the percentage of male instructors in period two over period one except HORT and PLPA that have significant increases. ALEC is the only department that had predominantly female instructors, but only in the second period (54%; note, percent males are reported). Percentage of male instructors is as large as 99% (in ENTO). Rank of instructors also vary between different departments and periods with no easily discernable patterns. In the first period for example, almost 92% of all instructors in ENTO are professors, while in the same period in ALEC only 19% of all instructors were professors. If significant, the percentage of instructors that graduated from non-AAU school decreased between the two periods. The percentage of instructors graduating from non-AAU schools ranged from 51% (HORT and BICH) to 84% (BAEN) in the first period and from 35% (ENTO) to 71% (WFSC) in the second period.

Student characteristics

Compared to the previous two groups of indicators, student related characteristics have more similarities in direction and magnitudes among the departments. All departments had decreases in the percent of male students between the first and second period except AGEC, which had no significant change. Percentages of male students, however, still show a broad range, from 40% in HORT to 80% in BAEN for the first period and 32% in ANSC to 78% in BAEN in the second period. Average SAT scores are significantly higher in all departments in the second period, except ALEC and SCSC. This may indicate that majority of the departments recruited better students over time. SAT scores visibly dropped in the last two semesters in almost all departments (Figure 8). The lower end of the range on average SAT scores changed little between the two periods, 526 (POSC) and 528 (ALEC), whereas the upper end has increased from 603 to 624 (both in BICH).

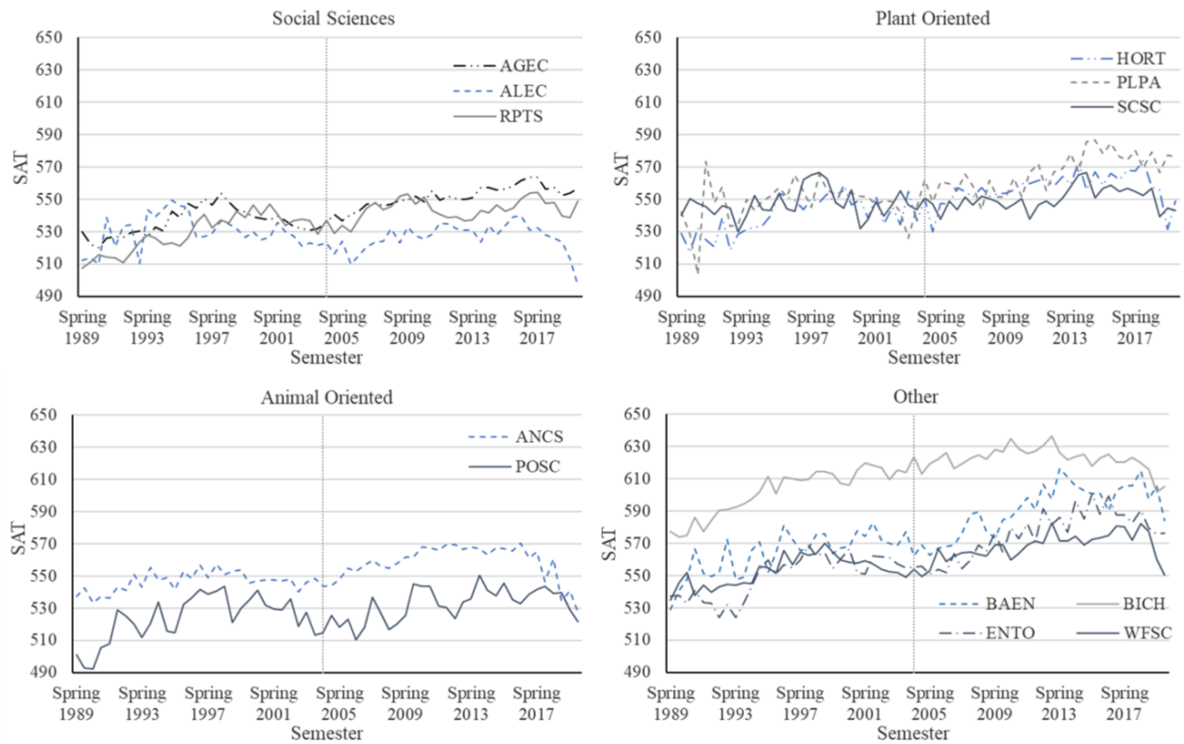


Figure 8. College of Agricultural and Life Sciences, Departments' Average SAT Scores by Semester from 1989 to 2019

In six departments (AGEC, ALEC, ANSC, SCSC, BAEN, and BICH), average student load significantly decreased and in three departments (RPTS, HORT, and WFSC) load increased. Several conflicting policy changes may impact student load. Texas A&M University gradually decreased the number of credits necessary to graduate from 140 to 120 between the mid 1980's and early 2000's. Currently, TAMU generally requires 120 credits to graduate. Students taking more than 150 credits are required to pay out-of-state tuition. The number of credits a student can take before having to pay out-of-state tuition decreased between 1999 and 2006. Students graduating with 123 or less credits may be eligible for a small tuition rebate. In Fall 2005, TAMU changed tuition from per credit to a set rate for students taking 12 plus credits. TAMU introduced flat versus variable rate tuition in 2014 where students entering the university can select a tuition plan for the next four years.

Average high school rank is 73 percentile or higher in all departments, meaning that in high school 73% or more of all students ranked below those students accepted to COALS (Figure 9). Four departments (AGEC, ALEC, BICH, and SCSC) had significant decreases in high school rank, whereas two departments (ANSC and ENTO) show increases between the two periods. Other departments' grouping which includes three STEM majors (ENTO, BAEN, and BICH) showed students high school rank increased in both time periods until the last couple of years. The share of no grade has either significantly decreased or has had not changed between the two periods for all departments. BICH (with over 5% of students receiving a no grades) had the largest percentage of no grades in both periods.

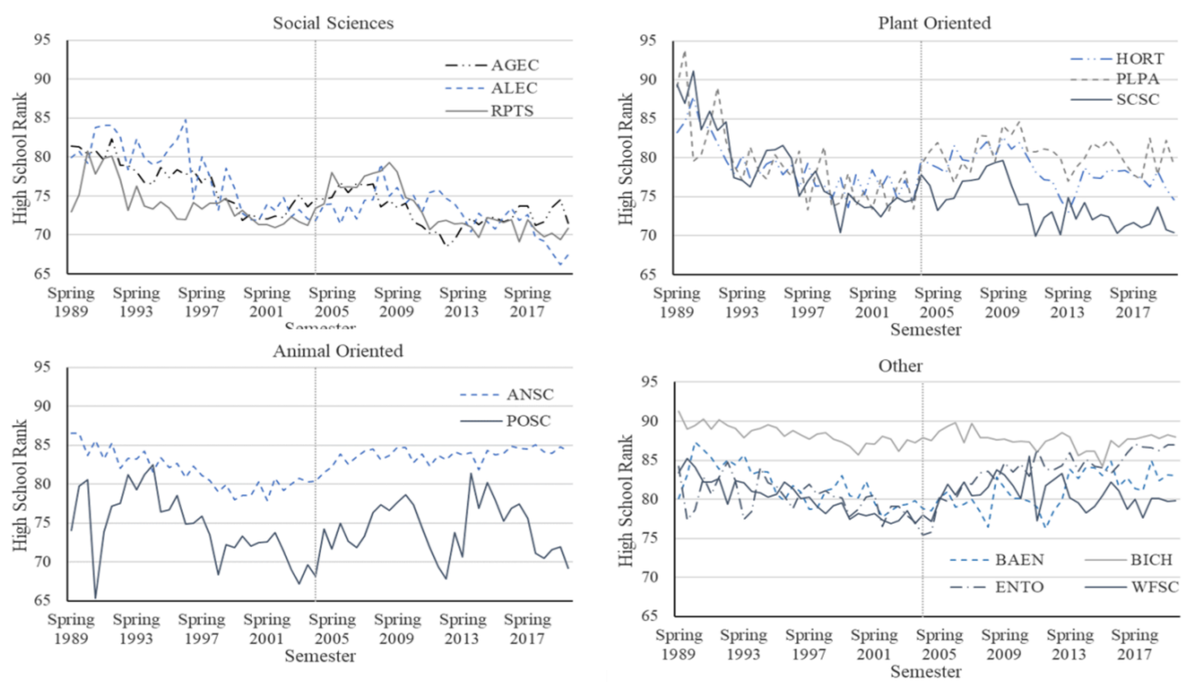


Figure 9. College of Agricultural and Life Sciences, Departments' Average High School Rank by Semester from 1989 to 2019

Model

Descriptive statistics show there are differences between departments within COALS. Further, different instructors have different teaching styles and may grade differently which may make the assumption of independence of observations invalid. To account for these differences, mixed effect models (Goldstein and Hoboken, 2011) are estimated individually for each department. The models consider instructor-specific characteristics to estimate the average grade in each class. Previous studies have also used mixed effect models in examining grading patterns (Kokkelenberg, Dillon, and Christy, 2008; Beenstock and Feldman, 2016; Hernández-Julián and Looney, 2016).

Mixed effect model estimation contains fixed and random components and consists of two levels. Level 1 equation is:

$$y_{ij} = \beta_{0j} + \sum_{m=1}^{k} \beta_m x_{mij} + \varepsilon_{ij} \quad (1)$$

where

y_{ij} - is the vector of observations (GPA) for the i^{th} class taught by the j^{th} instructor,

X_{imj} – is i^{th} class' m^{th} characteristic for j^{th} instructor (see variables in Table A3),

β_{0j} - is the vector of fixed effects, or mean GPA for the j^{th} instructor,

β_m – is the vector of coefficients for class characteristics, and

ε_{ij} - is the vector of residuals for the i^{th} class taught by the j^{th} instructor.

Level 2 equation is

$$\beta_{0j} = \gamma_{00} + u_{0j} \quad (2)$$

where

β_{0j} - is the vector of fixed effects or mean GPA for the j^{th} instructor,

γ_{00} - is fixed intercept across all groups, and

u_{0j} – is deviation of the j^{th} instructor from the fixed intercept.

Results

Given the number of models and variables, limited discussion is provided on inferences on the coefficients by department and period (See tables A5 and A6). As shown in the two tables, there are period and departmental differences in significances, signs, and magnitudes of some of the coefficients, but also many similarities.

Potential grade inflation

In the first period, one department (POSC) has a significant and negative coefficient associated with trend, whereas four departments (AGEC, SCSC, ANSC, and BAEN) have positive and significant coefficients associated with trend after controlling for the other characteristics. In the second period, negative significant trend coefficient is observed in BICH, while the number of departments with positive significant trend coefficients doubles (AGEC, ALEC, RPTS, HORT, PLPA, SCSC, POSC, and BAEN). It appears the increase in COALS grades reported in the first essay is caused by most of the departments experiencing increasing trend in grades especially in the second period, but not all departments.

Institution characteristics

Institutional characteristics appear to show few patterns concerning significance, signs, and magnitudes of the coefficients. Eleven of the 24 coefficients associated with morning classes are significant and all but two are negative. Only in PLPA (not significant), SCSC, and POSC are the sign and significance of this coefficient consistent between the two periods. Classes taught during morning hours (if significant) have lower GPAs than afternoon classes with the one exception, BICH in the second period. This is in line with Marbouti et al. (2018), who find that early morning and late Friday afternoon classes attendance and grades are lower than other meeting times. Classes meeting only once a week generally have higher grades. In period one, 10 coefficients are significant and negative for meeting two or more times a week. Only one department had a significant and positive coefficient (AGEC) for meeting two or more times a week. Differences between time periods are present. In period two, five departments had lower grades when meeting more than once a week, whereas four departments had higher grades when meeting more than once a week. For both periods, only ENTO and RPTS had no significant coefficients associated with number of classes per week.

Grades in upper and lower division classes for most of the departments are similar, that is the coefficients are insignificant. ANSC has higher grades in upper division classes for both periods. In contrast, SCSC had lower grades for both time periods in upper division classes. BAEN (period one) and ENTO (period two) grades are higher for upper division courses, whereas WFSC (period two) grades in upper division are lower than lower division courses. The number of students in the class is negatively correlated with grades for all departments and periods except for RPTS in period two where the coefficient is insignificant. This finding is in line with many studies who find student perform better in smaller class sizes (Kokkelenberg, Dillon, and Christy, 2008; Diette and Raghav, 2015; Nye, Hedges, and Konstantopoulos, 2001). If significant, courses with three or more credits result in lower grades relative to courses with one or two credits.

Instructors' characteristics

Instructor characteristics involving 138 coefficients (six are not considered, because they are drawn on a small number of observations) for all departments are only significant in 13 cases (six in the first and seven in the second period). In an ideal world, none of the variables in this group would affect grades. Only in SCSC "other instructors" is significant in both periods. In the second period, instructors' gender significantly affected grades in RPTS (males grading lower than females) and in ALEC (females grading lower than males). Assistant professors graded lower than professors in SCSC and ENTO in the first period. In PLPA for the second period, assistant professor grades are higher than professors. Associate professors grade significantly lower than professors only in ENTO in the second period. Graduating from an AAU or non-AAU school appears to have little to no effect on grading, especially in the second period. If significant, graduate students and other lecturers grade higher than professors, but this occurs only in five cases (period one ALEC and SCSC; period two RPTS, SCSC, and ANSC). Research suggests that one of the reasons for higher grades granted by visiting and adjunct faculty could be the expectation of higher student evaluations (Kezim, Pariseau, and Quinn, 2005; Sonner, 2000). These instructors are often hired on a term-by-term basis and higher student evaluations are more likely to result in their contract being extended. But this does not seem to be the general case in COALS.

Students' characteristics

Students' characteristics have more significant coefficients compared to instructors' characteristics. If significant, students' characteristics generally have similar inferences in all departments: decrease in percentage of male students, as well as increases in SAT score and high school rank have positive effects on GPAs, while increase in share of students with no grades has a negative impact on GPAs. Studies such as Voyer and Voyer (2014) and O'Dea et al. (2018) also find females tend to receive higher grades. In two of the four departments where student gender is insignificant in the second period, the percentage of female students is larger than males (ALEC and HORT).

In five departments, RPTS, PLPA, SCSC, POSC, and WFSC, for both periods, increasing SAT scores did not significantly increase GPAs. Although you would expect SATs to reflect students' ability, studies such as Haladyna, Nolen and Haas (1991) and Reames and Bradshaw (2009) support the idea that SAT scores have increased over time without a corresponding increase in student educational achievement. They claim this may be a result of public schools preparing students to take standardized tests or applying a more effective block scheduling. High school rank, reflecting student preparedness and motivation (Kobrin et al., 2008; Westrick et al., 2015), is insignificant in five departments in the first period but is only insignificant in three departments in the second period (RPTS, SCSC, and WFSC). As expected, the percentage of students receiving a no grade is generally associated with lower GPAs. This is most likely because more students tend to drop difficult courses compared to easier classes. Barker and Pomerantz (2000) state dropping a course may suggest poor performance and indicate responsible behavior by the student who is considering their academic future. Five coefficients are significant and positive and two are significant and negative for student load considering both periods. These findings weakly suggest motivated students and students with less free time do not procrastinate and organize their time more wisely. This may result in better study habits and higher grades. Previously discussed changes in university policies that may influence student load, may be leading to this characteristic being insignificant in many departments.

Conclusions and Discussions

Differences in grading patterns (specifically, class average GPAs for periods 1989-2004 and 1989-2019) among 12 departments within the College of Agriculture and Life Science (COALS) at Texas A&M University (TAMU) are examined through addressing two objectives. First objective is to determine if potential grade inflation has been occurring by department and if it differs overtime. A significant and positive coefficient for trend indicates potential for grade inflation, but it must be noted missing variables may be contributing to the trend coefficient. Potential grade inflation occurred in four departments in the first period (1989 – 2003). In the second period (2004 – 2019), the number of departments experiencing potential grade inflation doubled to eight out of the 12 departments. Three departments, AGEC, SCSC, and BAEN had potential grade inflation in both periods. In contrast, in each period only one department had potential grade deflation (POSC in period one and BICH in period two). POSC experienced grade decreases in the first period and increases in the second period.

Although not in the model, the change in number of departments experiencing potential grade inflation roughly corresponds to factors previous studies suggest as reasons for grade inflation including tuition and fee increases, increase in the use of teaching evaluations, and student generation. The second period roughly corresponds to the time when baby boomer (including younger baby boomers – generation Jones) were ending their student careers and millennials (generation born between 1980 and 1996) started attending college. By the end of period two, Generation Z started to enroll in college. Howe and Strauss (2000) mention millennials were raised by their parents to succeed. In addition, Curran and Hill's (2019) meta-analysis shows recent generations of college students feel more pressure to excel than students in 1990s. This need to excel could be one driving force behind students' complaints on grading and could foster grade inflation. Additional research on generation on grading patterns is warranted.

GPAs show a decline in both ALEC and BAEN around 2006, however, they crawl back up by the end of the second period. Discussion with the BAEN former department head indicated they attempted to increase rigor in their department. These observations imply grades are hard to reduce and / or maintain at lower levels.

The second objective is to examine factors influencing mean class GPA among different departments in COALS to provide information on factors causing these differences and explore if the influence has changed over time. Results show that there are differences in grading patterns among departments in COALS and even within the same department between time periods. It appears differences in GPAs are mainly driven by specifics of each department. This is in line with the first essay where significant departmental differences are reported. Departmental culture, subject matter, job market prospects, and student expectations may be some of the reasons for departmental differences. These differences may manifest themselves in the magnitude of the coefficients differing although sign and significance are the same. Although departmental differences may be the main driving force some differences are noted and discussed. Further, because of these differences, one must be careful in comparing students and their GPAs between majors; an unfortunate inference for employers and graduate school recruiters.

In terms of ranking from the largest to smallest GPA between periods, only three departments had a change of more than two places in this. BAEN changed from fourth to seventh in its ranking. As noted earlier, BAEN made a conscious attempt to add rigor to their program. No reason is found for the other two departments' change in ranking. WFSC went from seventh to 12th with a significant decrease in GPA between periods. PLPA with the largest increase in GPA between period went from 10th to sixth. The remaining discussion concentrates on period two as noted earlier this may be the most relevant period.

Weak evidence exists that supports previous studies' claims that differences exist between STEM and non-STEM designated departments. Seven of the eight non-STEM departments experienced potential grade inflation while two (PLPA and BAEN) of the four STEM designated departments (PLPA, BAEN, BICH, and ENTO) experienced potential grade inflation. As noted earlier, changes in BAEN grading may have more to do with changes in the department than STEM designation. One STEM department shows grade deflation (BICH) and one no change (ENTO) in GPA. Over time, grade dispersion among all departments reduced from a GPA range of 2.72 - 3.50 in 1989 to 3.05 - 3.52 in 2019, making it more difficult to differentiate students' abilities.

Issues remain what are causing the difference between departments. After controlling for instructors, characteristics associated with instructors are generally insignificant implying these characteristics are not the reason for differences. Signs and significance of student characteristics are similar among departments, but magnitudes vary. Simple correlation between estimated coefficients on high school rank and average school rank is 0.75. Such a moderate to strong correlation indicates the effect of preparation as given by high school rank is stronger in classes that have a higher average rank than classes with lower average rank. Correlations between the absolute value of the estimated coefficients and average values for student gender (0.38) and SAT scores (0.42) show a weak to moderate relationships. Although the effect of students' characteristics such as preparedness, motivation, and gender are similar, having a larger percentage of better prepared student, for example, has a larger impact (magnitude) on grades. More research is warranted on these relationships.

Institutional characteristics do not present as clear of a picture. Characteristics other than total students enrolled in a class and high credit show no consistent patterns. Correlation between estimated coefficients and average number of high credit classes is very weak to nonexistence at -0.16. Negative correlation between estimated coefficients and average number of students in a class shows an inverse moderate to significant relationship (-0.68). Although increasing the number of students decreases grades, it appears at some point adding additional students has less of an effect. This indicates the relationship between number of students and grades may be nonlinear. At some point increasing the number of students may have little to no effect on class GPA. Again, more research is necessary on this relationship.

Questions not addressed include should grade reform be undertaken and whether departments would be willing to consider grading reform. These are complex difficult questions involving issues such as enrollment, finance, and employment. Because administrators may not have a lot of control over individual instructors' grading standards, they may introduce the idea of "individual gain" (McGowen and Davis, 2022). Individual gain is a numeric value calculated based on the initial test and a final test at the end of the class that can be used to complement grades on students' transcripts. Such a numeric value, however, would be a confusing addition to transcripts especially until all universities adopt the idea.

Denning et al. (2022) show grade inflation has led to an increase in college graduation rates, a goal of the current presidential administration. Compared to education expenditures grade inflation may be a low-cost policy option to ensure higher graduation rates and earlier graduation. However, the long-term consequences of such a policy, such as decline in quality of college graduates or university image deterioration needs to be considered. Future research calculating costs and benefits that come with increasing grades. Benefits comprise of higher rates of completing college, which results in graduates who compete for better employment opportunities. Costs include lower preparedness of those graduates.

Given that this essay utilizes the same dataset as the preceding one, it is important to acknowledge that limitations encountered in the first essay are also applicable here.

POLICIES ENHANCING AGRICULTURAL VALUE CHAINS IN DEVELOPING COUNTRIES: AN EVIDENCE GAP MAP

Introduction

Agriculture, comprising approximately four percent of global gross domestic product (GDP), plays an important role in driving economic growth (World Bank, 2022). However, in many low, middle, and upper-middle income countries, where food security and malnutrition are significant issues, agriculture makes up more than a quarter of their GDP (World Bank, 2022). Agricultural growth plays an important role in poverty reduction and food security (Warr and Suphannachart, 2021; Abdelhedi and Zouari, 2020; Norton, 2004), and whose role in addressing these challenges has been widely acknowledged (Zadawa and Omran, 2020; Gero and Egbendewe, 2020). Promoting the growth of agriculture in low, middle, and upper-middle income nations is crucial not only for their own development but also for the well-being of the entire world, due to these countries producing nearly half of the global food supply (USDA Economic Research Service, 2017).

Agricultural Value Chains (AVC) are mechanisms that allow for the flow of products, knowledge, and financial payments possible from producers to final consumers. Strengthening AVCs to ensure the supply of food is especially important for developing countries where food security and malnutrition remain concerns. Agricultural growth is important not only for rural poverty alleviation, but also for reducing urban poverty (Norton, 2004). It has also been shown that participation in global AVCs increases GDP and employment in the agricultural sector (Lim, 2021). Supporting policies for creating enabling environments for AVC is one of the priorities of global development organizations, such as FAO, World Bank, and USDA (Clayton and Preston, 2003; Smyth, Phillips, and Kerr, 2016). The World Bank book on agribusiness and innovation systems in Africa mentions that developments in agriculture reduce global poverty directly through raising farmers' income and indirectly through decreasing food prices. In many developing countries the agricultural sector is the largest source of employment compared to other sectors (Larsen, Kim, and Theus, 2009). This suggests that enhancing employment opportunities within the agricultural sector might hold greater potential for poverty reduction than doing so in other sectors.

However, effective and efficient coordination of AVCs continue to remain an issue in low, middle, and upper-middle income countries due to capacity constraints and inefficiencies of agricultural policies (Trienekens, 2011; Protopop and Shanoyan, 2016). Value chains serve as mechanisms for the introduction of alternative organizational systems that allow more efficient production, processing, and marketing practices. An example is efficient warehousing and retailing systems in large supermarkets that utilize technologies for fast and efficient order refills (Liu et al., 2019). Thorough analysis of interventions in AVCs and the resulting outcomes reveals areas that are underexplored and may have a potential of a higher marginal contribution to agricultural efficiency, and stakeholders' profits, and society welfare.

To shed light on impacts of existing policy interventions, one of the USDA Food for Progress Learning Agenda questions focuses on identification of gaps in implemented policies along AVCs, as well as on revealing models of collaboration that are effective in enhancing AVCs (USDA, 2020). Specifically, the aim of the Food for Progress Learning Agenda on Trade Expansion and Agricultural Market Development is to "…identify relevant and timely research questions to inform evaluation and policy research in the area of expanding agricultural trade and markets" (USDA, 2020). Informed decision making in policy design is crucial, especially given limited resources for interventions. Policy interventions (any action, program or activity implemented by national or international authorities and non-state actors) at different stages of the AVC affect desired outcomes (yield, technology adoption, poverty reduction, etc.) differently with the complexity of the chain further increasing the difficulty to measure the impact of a specific intervention. It is important to evaluate policies and their impacts based on actual evidence (coming from real-world examples, rather than simulations or theoretic contributions) preferably collected from a wide range of applications and studies.

The objectives are collecting evidence on interventions in AVC and highlighting evidence gaps in the literature. Further, models of collaboration aimed at improved policy interventions in any stage of the AVC are discussed. Specific questions addressed are:

- What are the extent and characteristics of existing empirical evidence regarding policy interventions that enhance AVC and improve enabling environments and what are the evidence gaps in the literature?
- What are the extent and characteristics of existing empirical evidence regarding the effectiveness of models of collaboration among local and international actors, including donors, private sector partners, academic institutions, and Non-Government Organizations (NGOs) to support policy change?

Evidence Gap Map (EGM) (Snilstveit et al., 2017) approach is used to achieve the first objective by mapping existing knowledge through systematic screening of the literature on AVCs obtained by keyword and manual searches of electronic databases and organizational websites. EGM utilizes a systematic data collection approach to visually present information that is known or not known on a specific topic; therefore, allowing organizations and agencies to target gaps potentially leading to higher impact policy interventions. In addition, because the evidence is based on data coming directly from the field (excluding simulations and theoretic studies), the results may be more realistic and generalizable. The EGM developed combines evidence on interventions along AVC with resulting outcomes. Specifically, interventions are separated by stages of the value chain (input supply, production, harvest, and post-harvest, transport and storage, processing, marketing, as well as overall governance of the value chain) and mapped with corresponding outcomes (production outcomes, market reforms, product movement in space and time, risk management, and welfare). It is important to keep in mind that this method of data collection does not attend to significance or magnitude of coefficients, and only records whether the outcome was considered by included studies' authors. This implies that the intervention /outcome pair is included in EGM even if the intervention was not a significant driver of the outcome.

Second question is addressed via thematic synthesis of main themes emerging from the literature. Findings are based on frequencies of each theme. Specifically, models of stakeholders' collaboration are identified and then grouped based on different characteristics. Further, groups of collaboration models are matched with specific stages of the AVC and the policies they intend to promote.

This study is important to both policy makers and researchers. Based on the results of the EGM policy makers may want to reallocate funding to underexplored areas and will be able to identify interventions that are more likely to result in the desired outcomes. Likewise, researchers can use the information to identify areas where the research is either weak or ambiguous to further advance knowledge in those domains.

Scope

This EGM presents a framework of 17 interventions and 28 outcomes. These categories are informed by the conceptual framework defined in the next section and are adjusted based on the literature.

Conceptual framework

Theory of change (Weiss, 1995) is a process of documenting how and why an intervention is supposed to work. A theory of change diagram represents description of the strategies and actions that facilitate change and achieve outcomes. Using this framework, one can describe the causal relationship between policy interventions and their effects at different stages of the value chain. The following discussion outlines the key elements of the theory of change for interventions in AVCs. Causal paths in AVCs are complex interactions involving many stakeholders and institutions arrangements. Further, the nature of interventions depends on the stage of the value chain including input supply, crop production, harvesting, post-harvest treatments, transportation, storage, processing, and marketing. Norton (2016) discusses an AVC model which highlights intervention strategies that can be used in different stages of the value chain to achieve different outcomes. Policy interventions are implemented through local and international actors, government bodies, NGOs, and actors in the value chain itself. These interventions are specific to stages. For example, interventions at the input stage might provide seed and equipment at the time of planting and chemicals supply for fertilization and diseases control. Output of such interventions at the input stage may be increased production and / or farm incomes. Interventions in production, harvest, and post-harvest stages may allow for the possibility of increased participation in markets. Similarly, interventions at the processing stage may improve product quality which leads to higher prices for producers resulting in increased farm incomes. Therefore, understanding linkages and interactions within AVCs helps identify key areas of interventions that lead to the desired outcomes. These linkages are derived from Norton (2016) and presented in the theory of change diagram in Figure 10.

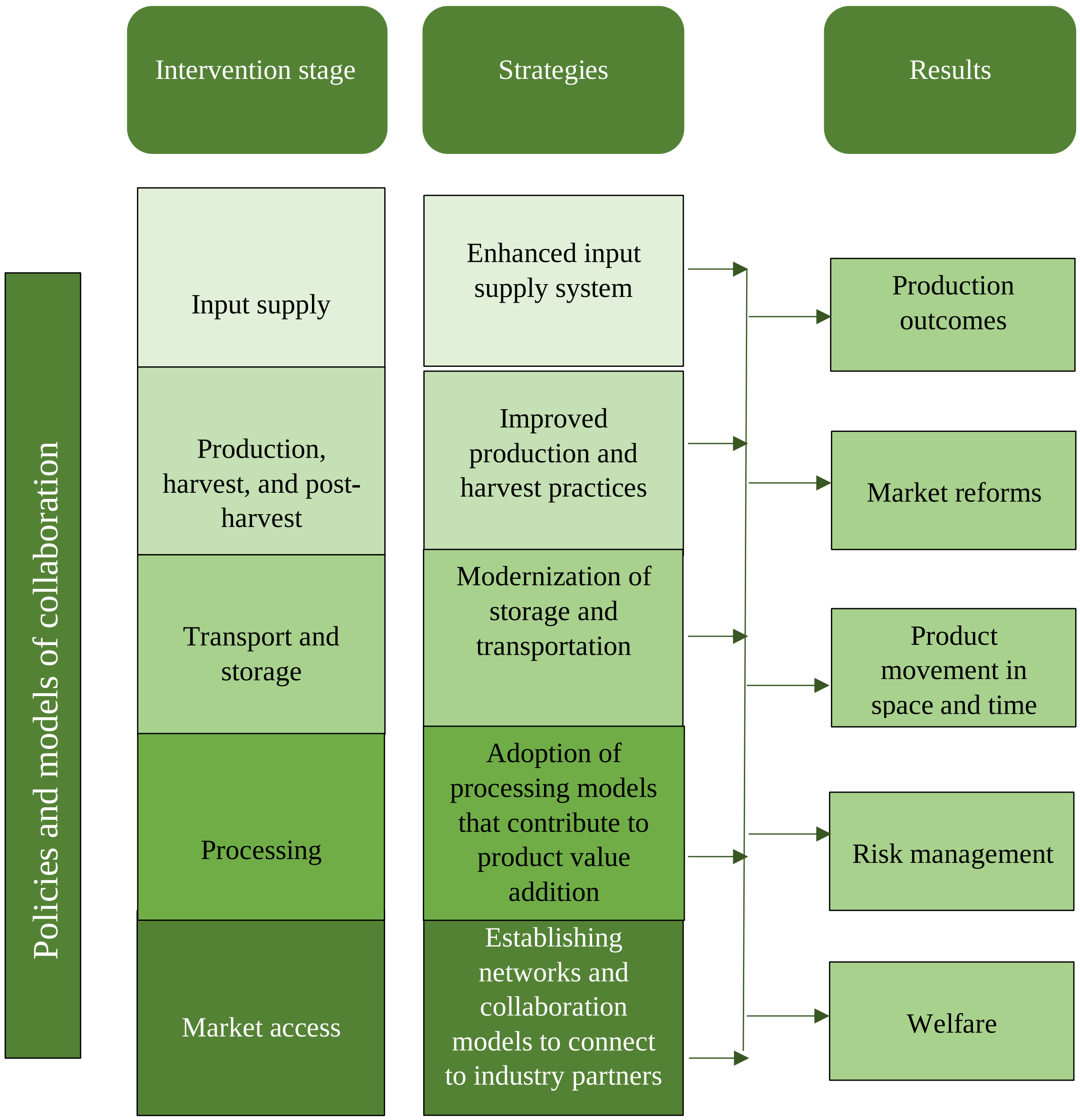


Figure 10. Scope of the EGM: Theory of Change Diagram

Theory of change diagram presents AVC stages and strategies that can lead to improvements of AVC at each stage. The last column contains potential results obtained after the intervention takes place. For example, improved production and harvest practices implemented in production, harvesting and post-harvest stage of the value chain can lead to improvements related to market expansion or product movement in space and time.

Although not presented in the figure, this approach requires assumptions about interventions scope, design, implementation process, and causal impact on resulting outcomes (Jones and Rosenberg, 2018). For example, it is assumed that the interventions were delivered as intended, were readily adopted by farmers, industry, and consumers without any modifications, and were actually completed.

Description of interventions

Agricultural policies are relatively permanent public decisions aimed to promote the development of institutions to fulfill the requirements of a rural economy, and influence prices received by farmers and their access to resources, inputs, and markets. Policy interventions can be implemented through new legislation, decrees, and public investments and programs. Modern agricultural policies are classified in three broad classes: pricing, resource, and access policies (Norton, 2004). Pricing policies are determined by the macroeconomic policies of the countries and affect the production decision of the producers by changing the relative prices of the goods and services produced. Examples of such policies are price controls, farm support prices, domestic subsidies, export incentive policies, exchange rate policies, tariffs, import quotas, food regulation and standards, free trade / trade liberalization policies. Resource policies are aimed to provide access and / or improve management of resources. This category includes land tenure, soil conservation, water use / irrigation, chemical use, and water resource policies. Finally, access policies make producers competitive in the production process and improve the overall efficiency and capacity of the AVC. These interventions improve producers' access to markets (both input and output markets) and technology. Examples of access policies include agriculture markets and storage facilities, road and transportation, agriculture-finance and credit, technology adoption, extension services, agriculture insurance, and agricultural cooperatives policies. Interventions through these policies directly affect a specific stage of the AVC (Norton 2016).

In addition to focusing on the policies' impact on different actors, value chain approach also examines interactions among the different actors. These interactions and relationships involve the governance of value chains. Governance refers to the inter-firm relationships and institutional mechanisms through which non-market coordination of activities in the chain is achieved (Humphrey and Schmitz, 2001). The degree of relationship and power between actors determines what, how, who, when, where, and why of production and processing decisions. Keeping in mind issues of poverty and high rates of female participation in agriculture in low, middle, and upper-middle income countries, AVCs should also be inclusive to incorporate needs of socially diverse segments of the population.

To summarize the domain of AVC interventions, interventions can be either related to value chain governance and inclusion and be implemented broadly along the entire value chain, or can be grouped by the stages of value chain as follows:

- Input supply stage,
- Production, harvest, and post-harvest stage,
- Transport and storage stage,
- Processing stage, and
- Marketing stage.

Using the theoretical background presented, the classical agricultural value chain is modified in this study to add governance and inclusion to intervention groups already presented in Figure 10.

Description of outcomes

Performance and effectiveness of interventions is measured through a set of outcomes that can be both related to improvements for value chain actors and for society (Kaplinsky and Morris, 2000). Muflikh, Smith, and Aziz (2021) mention economic related outcomes (both financial and non-financial), as well as social and environmental outcomes. Outcomes are categorized into five groups (Table A7). The first group relates to the production process; examples are yield (Mason, Wineman, and Tembo, 2020; Aye, 2017) and adoption of new technology (Wong et al., 2020). Outcomes relating to market reforms comprise the second group; an example is increased profits from sales (Do and Park, 2018; Rashid et al., 2013). The third group is product movement in time and space. An example is product procurement services (Aye, 2017; Van Campenhout, Minten, and Swinnen, 2021). Risk management (a group of external outcomes that usually are not controlled by value chain actors) is the fourth group. Examples are risk related to price uncertainty or contract risk (Shobana and Gandhimathi, 2015; Pavez et al., 2019). Finally, societal welfare forms the fifth category. Examples are changes in family savings and expenditure (Aye, 2017) or enhanced food security (Muyombano and Espling, 2020).

Studies of interest

To be included in the EGM, studies had to be published after 1999, written English, and discuss AVCs whose production component is in low, middle, or upper-middle income countries. Inclusion and exclusion of a study were implemented in accordance with Population, Interventions, Comparison, Outcomes and Study Design (PICOS) framework (Richardson et al., 1995). Each component is briefly discussed with a detailed description presented in Appendix B1.

Population: In included studies AVC stakeholders had to be input suppliers, farmers, producer groups, traders, processing companies, retailers / wholesalers, or consumers in low, middle, and upper-middle income countries according to World Bank (2019) classification. Indirect participants such as governments, local and international NGOs, and development agencies, were also included. Studies related to urban agriculture, as well as wildlife use, forest extraction, and wild fisheries were not included.

Interventions: A broad range of interventions that can occur at any stage of the value chain were included. Interventions included are policies adopted at the government level, as well as practices adopted by the stakeholders. Examples of formal policies are trade liberalization or protectionism, resource conservation and land tenure, provision of agricultural subsidies and credit, promotion of technology adoption, extension activities, or market access. Practices can be adopted by stakeholders with or without direct participation of governments. Examples are contract farming, certifications, agriculture information services, and other market mechanisms of "self" adjustment. Evaluation of the normal course of action of an organization or policies that had unintended effect on value chain are not included.

Although intervention categories were pre-defined using intervention categories suggested by the literature, the data extraction process refined intervention categories. A full list of intervention categories and corresponding supply chain stages are presented in the Table A8.

Comparison: Included studies had to involve comparison of policies and practices for an intervention. This is an essential part of designing a study to test treatments. Comparisons include:

- activities under one type of intervention vs another type of intervention,
- activities under intervention vs. doing nothing, and
- early-vs-late comparison in the implementation of an intervention.

Studies presenting only a diagnosis of a value chain and historical analysis without a comparison are not included.

Outcomes: Although suggestive outcome categories were identified before the literature search, some categories were amended by the results from included studies. This resulted in finalized outcome categories as presented in Table A7.

Study Design: Types of studies included are both impact evaluations (single paper that contains "intervention- outcome" relationship) and systematic reviews (groups of papers summarized into one paper). Included studies had to exhibit evidence of causality. In other words, the design of the study had to be such that any, if any, change in outcome variables had to be attributed to the intervention (program or policy) implemented. This implies that the study with appropriate design was included even if the difference between control and treatment was not significant. For studies providing quantitative evidence, the evaluations included used randomized designs, quasi-RCT, natural experiments, or methods to identify causation among self-elected groups (pre-and-post test data with comparison, multiple pre- and post- test data without comparison, cross-sectional with comparison, and post-test studies using instrumental variables). Studies providing qualitative evidence utilized discourse analysis techniques, thematic analysis techniques, grounded theory, phenomenological studies and ethnographic methods, or any combination of them (triangulation approaches). Studies using any combination of the quantitative and qualitative evidence listed above (mixed methods) were also included. Simulation studies or hypothetical experiments, narrative analysis, policy announcements, and summaries or overviews with no detailed reports on the impact evaluation were excluded.

Methods

Search strategy

Literature search was implemented through keyword search from major databases specialized on agricultural and economic publications. Agricola, Business Source Ultimate, Academic Search Ultimate, Cab Abstracts, and EconLit with Full Text databases were systematically[1] searched in April 2022. In addition, searches of fifteen development organization websites[2] were conducted.

Study screening and data extraction

Two reviewers independently screened each study using Covidence platform (covidence.org, 2022). If the two reviewers disagreed, the disagreements were resolved by a team leader. General information, as well as context and study design related evidence were extracted from each study. General information includes study ID number, authors, country, publication date, and category. Context related information includes intervention description, commodities affected,

[1] Systematic literature search refers to organized search process in a structured and preplanned manner. It demands careful consideration of search terms, selection of databases, and choice of search methods. Systematic literature search provides a greater chance of avoiding disparities and bias and enables to identify gaps in the existing research. (Aarhus University Library, 2022).

[2] UN Digital Library, 3ie, United Nations Development Program, Food and Agriculture Organization, International Fund for Agricultural Development, World Bank (including World Bank Data Catalog, World Bank Policy Research, and Independent Evaluation Group), Agriculture and Food Organization, United States Agency of International Development, Consultative Group on International Agricultural Research, Economic Research Services of the United States Department of Agriculture, Agricultural Research Service of the United States Department of Agriculture, Inter-American Development Bank, African Development Bank Group (including Independent Development Evaluation of African Development Bank), AgEcon search, EconPapers of Research Papers in Economics.

and outcomes reported (the outcomes were included regardless of significance level and the magnitude of change). Study design includes methods used in included studies. As in the screening stage, data was extracted independently by at least two reviewers. Overall, these procedures help ensure that selection of studies and data extraction closely followed inclusion / exclusion criteria.

Quality Assessment of Systematic Reviews

Because systematic reviews combine findings from many studies, methods used in the review for combining those studies may limit bias and improve reliability and accuracy of conclusions. This means there is a need for additional quality screening for systematic reviews. Systematic reviews were assessed for the quality using Cochrane Handbook for Systematic Reviews of Interventions (Higgins et al., 2022). Quality assessment is based on the compliance with the following criteria:

- requirements for databases searched (including grey literature, published outside of the traditional commercial or academic publishing and distribution channels),
- criteria used for deciding which studies to include in the review (including a question on compliance with PICOS criteria and a question on appropriate timeline for selected studies)
- specifications for screening and data extraction (independently done by at least two reviewers),
- requirements for quality of the selected studies,
- clear description of methods used and their appropriateness for a given task,

- conditions for combining studies, when applicable (including a question addressing heterogeneity concerns, a question on assigning proper weights, as well as a question on unit of analysis), and
- requirements for reporting main findings from included studies (including a question on summary of included studies, a question on separate analysis for studies with different risk of bias, and a question on factors that might explain differences in the results of included studies)

Each reviewer evaluated each requirement using a three-point Likert scale where a "three" indicated the requirements are satisfied, a "two" indicated requirements are partially satisfied, and a "one" indicated requirements are not satisfied. Scores for each point were summed to provide an aid in classifying the reviews into high, medium, or low-quality groups. Systematic reviews with scores from 14 to 21 were classified into low quality group, from 22 to 28 into medium quality group, and from 29 to 36 into high quality group. Seven high quality, nine medium quality, and eleven low quality systematic reviews were identified.

Evidence base

After removing 322 duplicates, 6,673 records were identified by title and abstract screening. One thousand sixty-nine studies were selected for full text screening with 193 studies being included for data extraction using the Covidence platform. Common reasons for exclusion are absence of robust study design, absence of comparison (before and after intervention), and study language other than English. Study screening and data extraction were implemented from June 2022 to January 2023. Details on numbers of studies are presented in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PICOS) (Moher et al., 2009) in Figure 11.

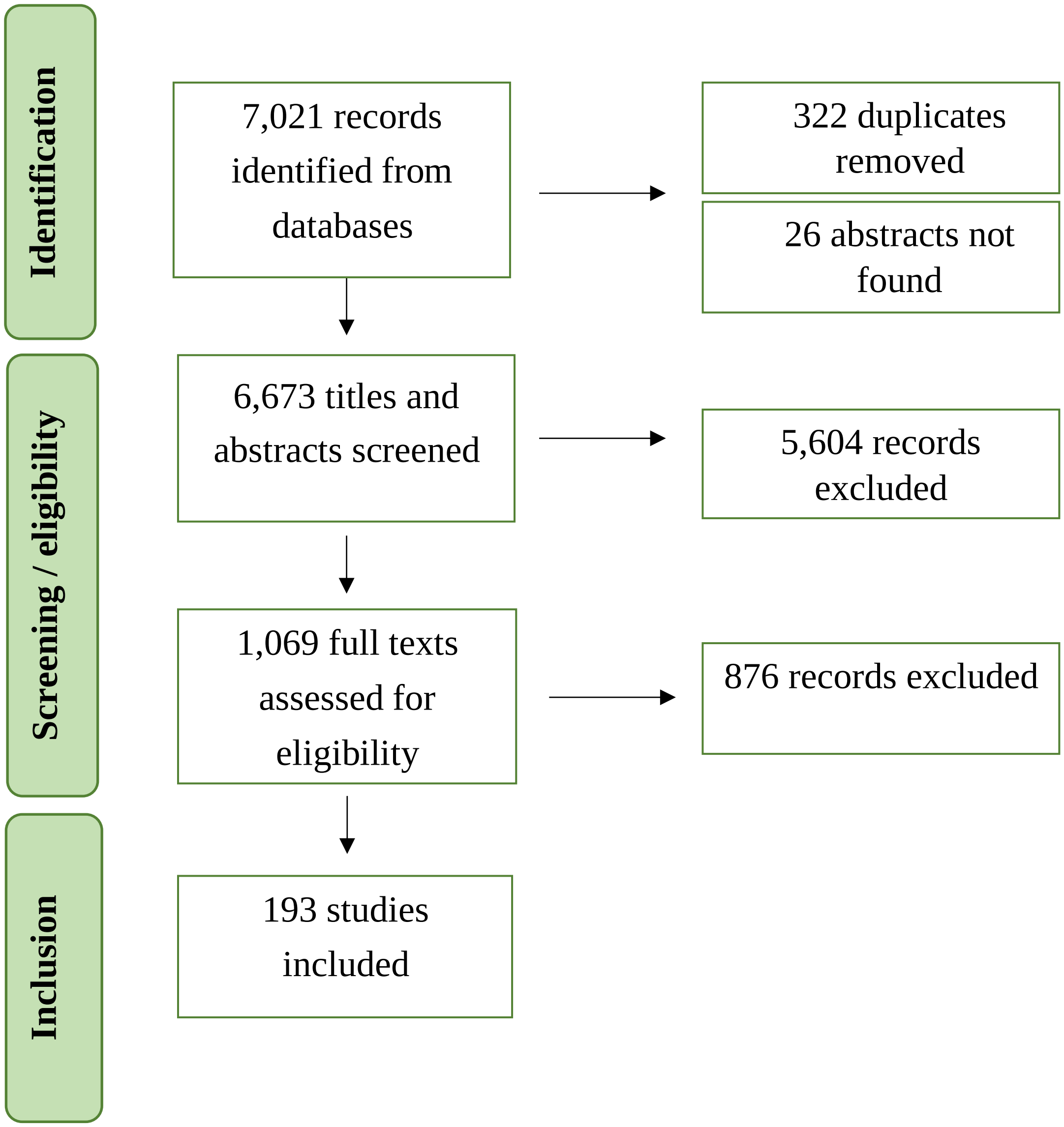


Figure 11. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Diagram

A listing of the 193 included studies is presented in Appendix B2. Data extraction questionnaires used for both impact evaluations and systematic reviews are included in Appendices B3, B4, and B5.

Findings

The first question of this study (impact of policy intervention) is answered utilizing data from both impact evaluations and systematic reviews, while the second question (effectiveness of models of collaboration to support policy change) is answered through studies discussing collaboration models. Some studies included information for both the first and the second questions. These studies are included in number counts for both impact evaluations and the collaboration models. The number of studies included for data extraction by type (impact evaluations, collaboration models and systematic reviews) are presented in Figure 12. There are clear differences between the nature of evidence regarding the first and the second questions answered by this study. The first question utilizes mostly quantitative methods, while the second one is primarily based on qualitative evidence.

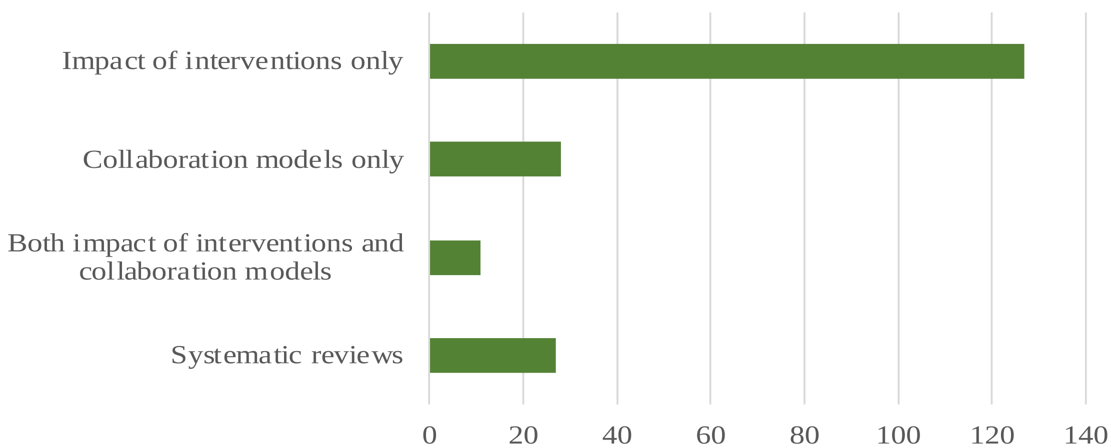


Figure 12. Number of Studies Used for Data Extraction by the Type of Study

About 75% of all included studies were published as journal articles. Next largest category was unpublished reports (11%). The remaining 14% were split between dissertations / theses, book chapters, or conference presentations. More than half of all systematic reviews were available through government or donor organizations websites. Only a very few publications come from the early 2000s, however the numbers increase after 2013 (Figure 13). The small number of publications from 2022 is explained by the search being implemented in Spring 2022; only studies published early in 2022 were screened.

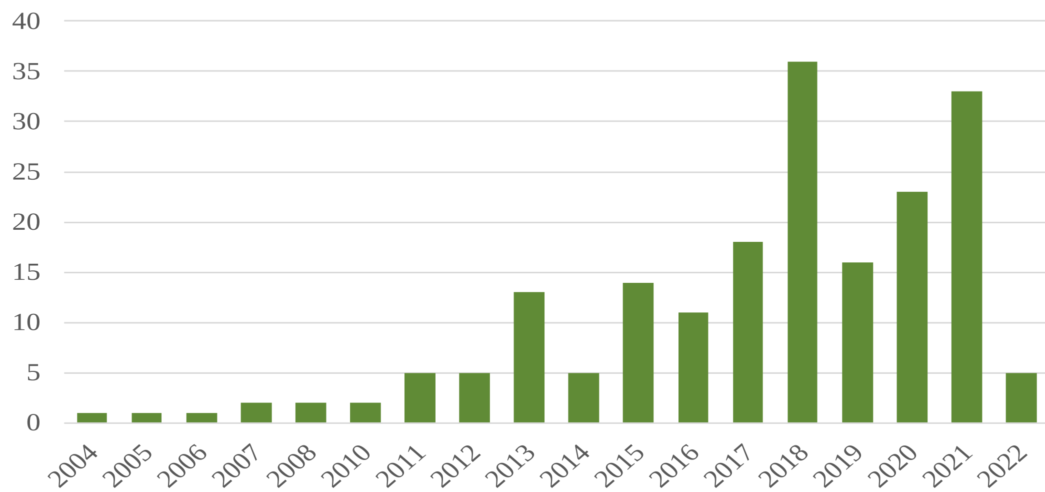
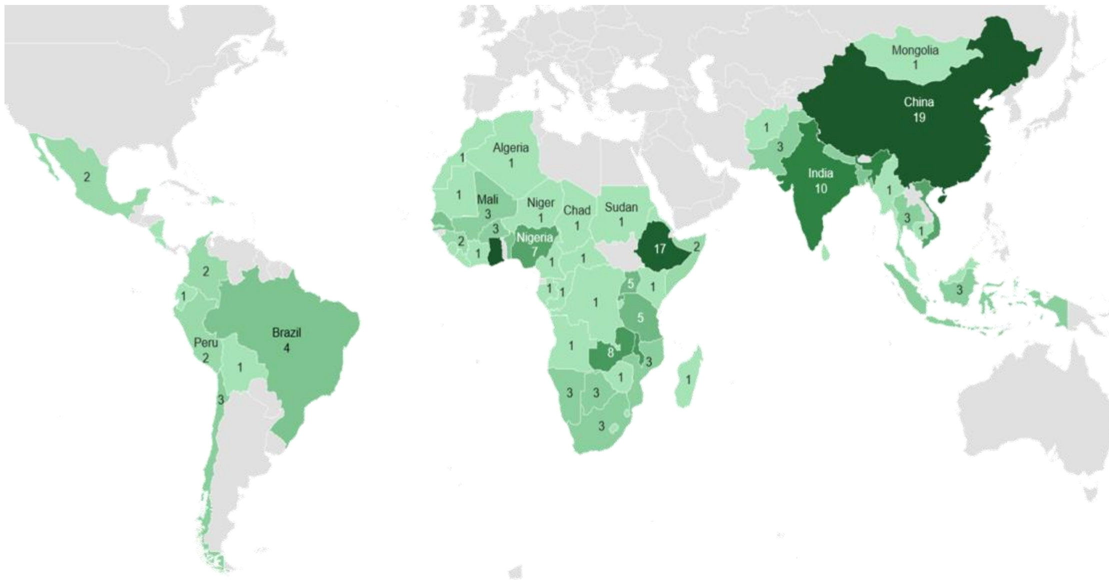


Figure 13. **Distribution of Included Impact Evaluations and Systematic Reviews by Year**

Map of Studies

The number of studies discussing impacts of interventions and collaboration models by country is given in Figure 14. The largest number of studies were conducted in Ghana, China, Ethiopia, and India (20, 19, 17, and 10 studies). Studies were conducted in all but four African countries. Fewer studies were conducted in the Americas and Southeast Asia.



Note: systematic reviews are not mapped due to the large numbers of studies included in each review.

Figure 14. Geographic Scope of Impact Evaluations and Collaboration Models

Impact Evaluations

Policy Interventions: Agricultural produce impacted by interventions were reported only for impact evaluations. Because studies were selected from low, middle, and upper- middle income countries, a slight majority (52%) of the crops impacted by interventions were staple crops (Figure 15). Staple crops include cereals, legumes, and roots and tubers. The next largest category was vegetables (16%), followed by fruits (9%), animal production, including meat and dairy farming (9%), and coffee, cocoa, and tea (9%).

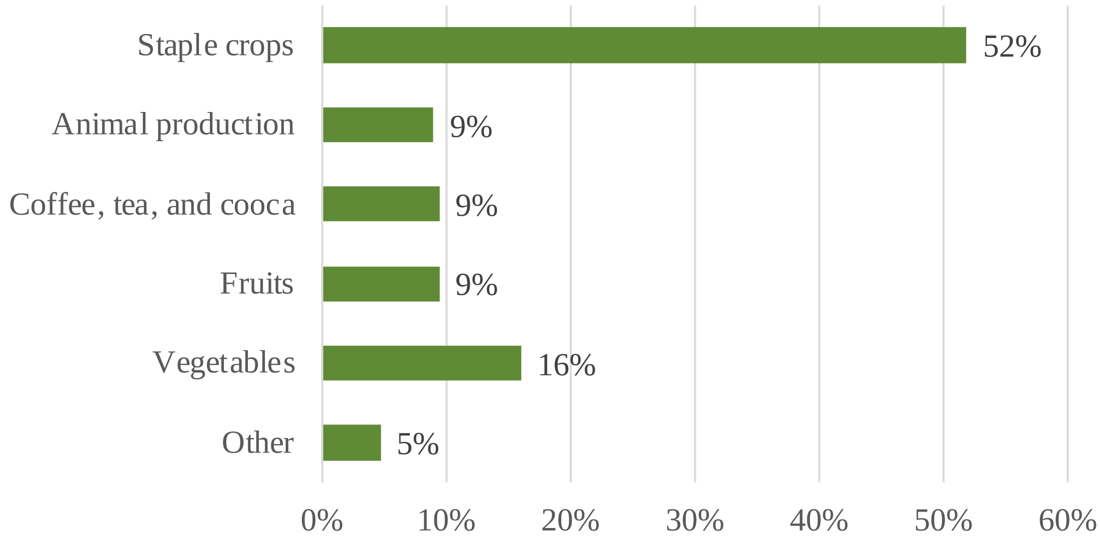


Figure 15. Agricultural Activity Directly Impacted by Policy Interventions

Note: some studies are not represented in the figure, because they did not report impacted activity, or the intervention was not related to a specific activity. Examples are trade policies or subsidized interest rates for agricultural loans.

Implemented policies mainly benefited smallholder farmers. In over 63% of studies these policies were initiated by governments - international, national, or local. Only 20% of the policies were initiated by stakeholders such as farmers, cooperatives, processors, and retailers. This implies governments play a significant role in shaping policies that affect the agricultural sector. Government's large percentage of involvement is partially explained by the fact that a considerable proportion of interventions carried out through international development organization were implemented through collaboration with governments. Out of the 15 cases where international NGOs or development organizations were engaged in policy implementation, only one did not include partnership with the government.

Nearly 80% of impact evaluation studies used quantitative methods including propensity score matching (PSM), randomized control trial (RCT), other quasi-experimental methods (50 studies), instrumental variables (17), difference-in-difference (12 studies), or models comparing intervention versus no intervention using Tobit or Double hurdle models (14 studies). Nearly half of the studies use multiple models.

Evidence on collaboration models

Collaboration models aimed at policy adoption were discussed in 39 studies. Multiple organizations worked together to implement policy modifications. Farmers or farmer groups as collaborators were mentioned in 85% of cases. Various industry stakeholders such as processors, traders, and intermediaries involved in importing or exporting were mentioned in 74% of cases. Governments were engaged in 46% of cases. International and development organizations were mainly involved through partnership with governments and national institutions. Although most of the collaborations had positive impacts on policy adoption, there were some studies that did not report clear impacts. The only collaboration with a clear negative impact involved collaboration between government and farmers that was funded by an international development organization. This collaboration failed due to inadequate interaction between the formal seed distribution agents (state extension services) and the private sector (local seed distributors and dealers) (Okry et al., 2011).

Factors responsible for effectiveness of the collaboration model (drivers of collaboration) were profits from collaboration, government support, external factors (international market developments, such as increase in demand), and multiple drivers resulting from full scale sector development (Figure 16). The main driver of collaborations was stakeholders' anticipation of potential profits.

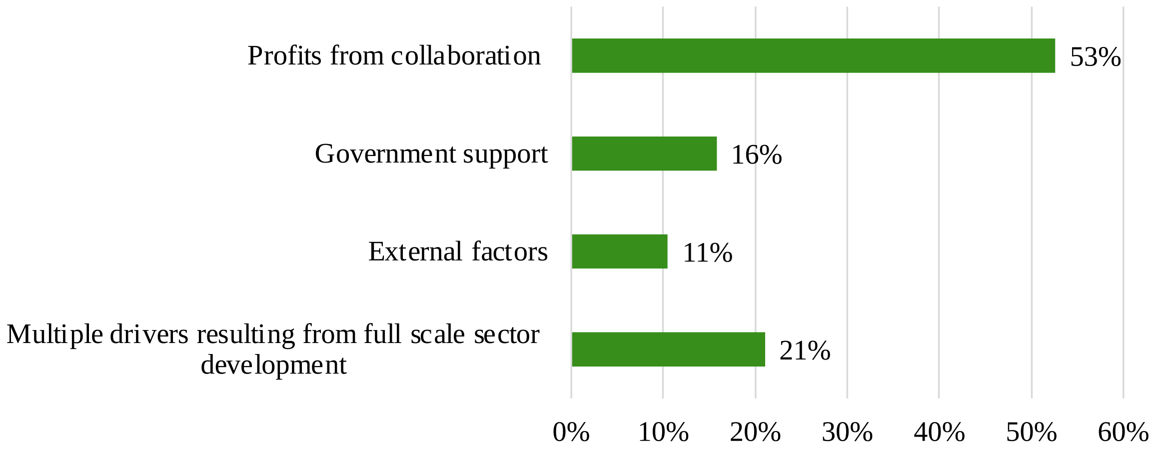


Figure 16. Drivers of Collaboration

Collaboration models were generally designed to examine potential policy change(s) which would enhance all stages of the value chain. Distribution of collaboration models along the stages of the value chain is presented in Figure 17. The largest share of the models was aimed at improved markets (38%) followed by governance and inclusion (23%).

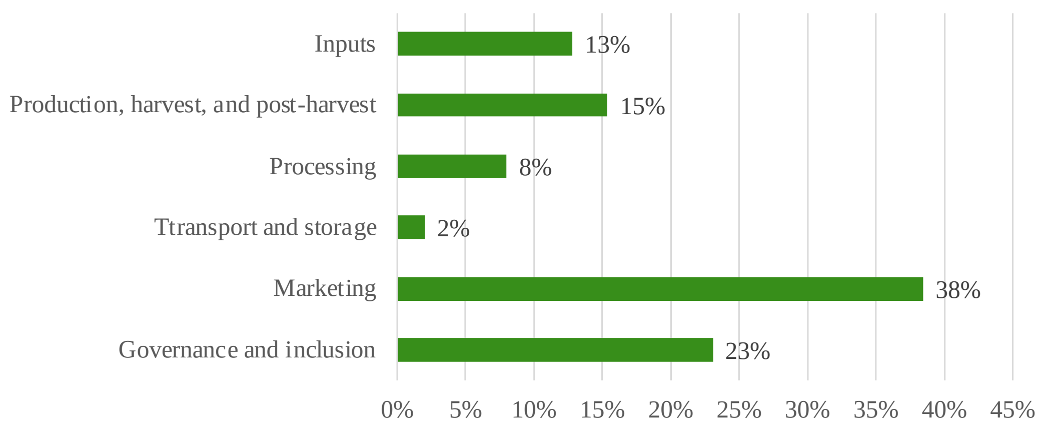


Figure 17. The Stages of the Value Chain Intended to be Improved

Systematic Reviews

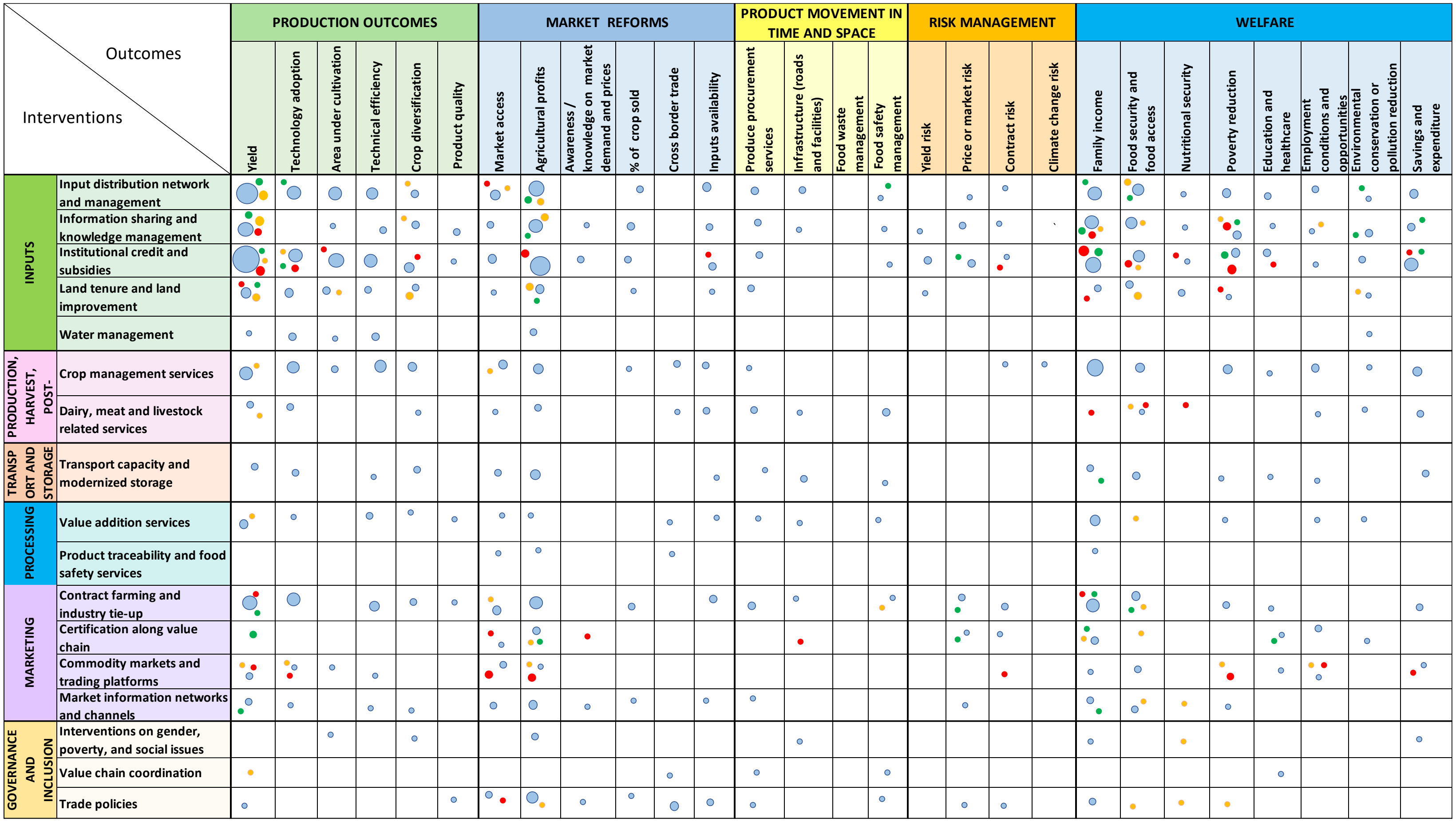
Interventions in the 27 systematic reviews were mainly centered around input supply (51%) and marketing (30%) stages of the value chain. These two stages also received the highest share of high-quality systematic reviews. There were no high-quality systematic reviews discussing interventions in production, harvest and post-harvest or governance and inclusion stages. In terms of outcomes, emphasis of systematic reviews was on outcomes related to production, welfare, and market reforms. Only two systematic reviews reported outcomes in product movement in time and space.

Seventeen systematic reviews used descriptive methods to summarize findings from the studies. Only ten used meta-analysis or meta-regression analysis to summarize the studies' findings. Most of the systematic reviews did not implement quality assessment of included studies. Language bias (a common bias for systematic reviews) was observed in the study selection stage, as only seven systematic reviews incorporated papers published in multiple languages. Furthermore, during the screening stage, nearly half of the reviews lacked independent screening by at least two reviewers, indicating a reviewer bias.

Gaps in AVC policy

About 85% of the interventions targeted only one group of stakeholders (usually farmers), while the remaining 15% had a broader scope, targeting multiple links or the entire value chain. Most interventions were at the input supply stage of AVC (Figure 18). Specifically, the largest number of studies were exploring institutional credit and subsidies. Second and third largest categories were input distribution networks and information sharing and knowledge management related to input supply. In terms of the number of studies conducted, following studies on input supply are studies on marketing, then production, harvest, and post-harvest stages. The most studied sub-categories in these two stages are contract farming and crop management services. Transportation and storage, processing, and governance and inclusion were not as frequently mentioned.

Common outcome categories were production, market reforms, and welfare, followed by product movement in time and space and risk management. Among outcomes the most frequent sub-categories are impact on yield, agricultural profit, and family income. The most populated intervention / outcome cells are in the intersection of input supply and production outcomes. Empty spaces in the figure signify intervention / outcome categories that were not discussed by previous studies. Risk management is the lease populated outcome category. Taking the above results together, agricultural policies and programs in low, middle, and upper-middle income countries are centered around decreasing poverty or increasing production, rather than on risk management. A sub-category that was never mentioned as an outcome is reduction in food waste. Food waste usually occurs at the retail and consumer levels when food is thrown away because of poor stock management or passed the expiration date. This type of loss is perceived to occur more frequently in high income countries (Rezaei and Liu, 2017). However, evidence indicates that even in countries that do not fall under the category of high-income nations, approximately 30% of food may still be wasted during the retail and consumption phase (Oelofse and Nahman, 2013). There is a clear gap in exploration of this topic, which provides opportunities for future research.



Note: An evidence gap map displays a matrix of interventions and outcomes combinations, indicating which interventions described in literature have resulted in which specific outcomes. The size of bubble increases based on the number of studies exploring the specific intervention / outcome combination. The same study may be included in more than one cell.

Blue bubbles represent impact evaluations, green bubbles are high quality systematic reviews, yellow bubbles are medium quality systematic reviews, and red bubbles are low quality systematic reviews.

Figure 18. Evidence Gap Map of Policies Enhancing Agricultural Value Chain

Frequencies of interventions and outcomes are presented in stacked charts in Figures 19 and 20 to help facilitate the presentation of the EGM. Interventions in input supply stage of the value chain were well represented in the literature. The least studied intervention categories are processing and governance and inclusion. The remaining stages had some moderate coverage. Only one systematic review explored interventions in processing and transport and storage stages, which is a direct result of there being only a few impact evaluations existing on these topics (Figure 19).

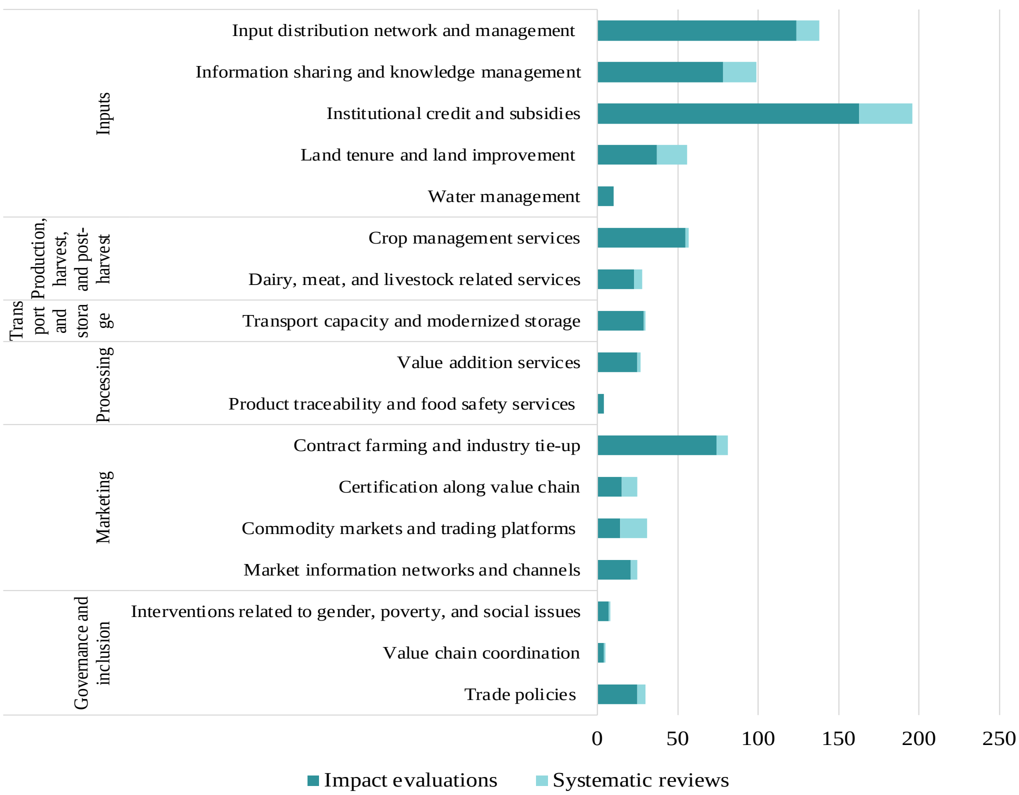


Figure 19. Number of Impact Evaluations and Systematic Reviews by Intervention Categories

As a result of value chain interventions, the most frequent outcome recorded was yield, followed by agricultural profit and family income (Figure 20). The least occurring outcomes among impact evaluations were those included in the risk management group (n = 27), while for systematic reviews least outcomes recorded were those in product movement in space and time group (n = 3). Reduction in food waste as an outcome category was not mentioned in any study, while climate change risk was mentioned only in one study. Evidence in existing studies did not prioritize these outcomes while discussing policies enhancing value chains.

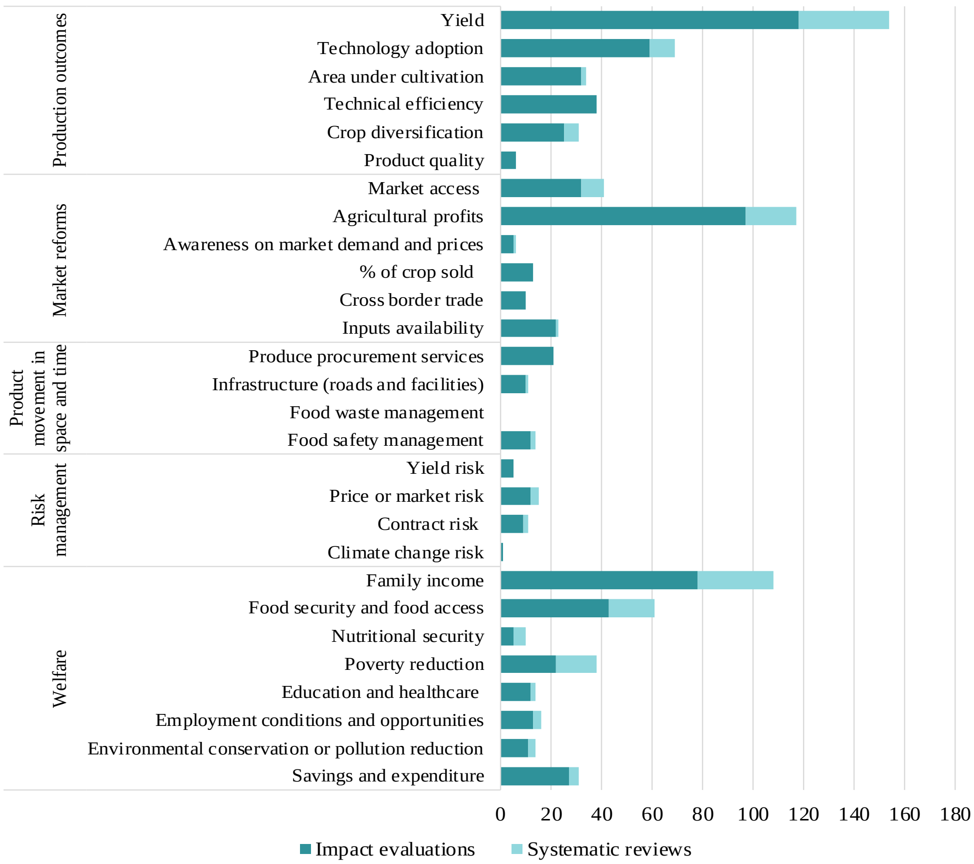


Figure 20. Number of Impact Evaluations and Systematic Reviews by Outcome Categories

Results by interventions: The frequencies of interventions with achievement of certain outcomes are given in the heatmap in Figure 21. Input stage interventions affect yield more than any other outcome category. Among the various outcome categories, yield is the most likely to be mentioned as outcome by the following interventions:

- Institutional credit and subsidies,
- Input distribution network and management,
- Information sharing and knowledge management,
- Contract farming and industry tie-up, and
- Land tenure and land improvement.

The next most frequent category of outcomes is agricultural profits followed by family income. It is important because since the main group of stakeholders in this study are smallholder farmers, agricultural profits and family income may often be hard to separate or differentiate from each other. Therefore, related outcomes to these two combined categories will more likely occur because of interventions in the following categories:

- Crop management services,
- Certification along value chain,
- Trade policies,
- Market information networks and channels,
- Transport capacity, and modernized storage,
- Value addition services, and
- Interventions related to gender, poverty, and social issues.

Compared to any other category, agricultural profits were mentioned as outcome more frequently as a result of interventions in the governance and marketing stage of the value chain. Interventions to promote commodity markets and trading platforms, market information networks and channels, as well as activities aimed at value addition services are more frequently examined as having outcomes related to market access and increased profits (indicated by dark green areas in the intersection of the mentioned categories).

		PRODUCTION OUTCOMES						MARKET REFORMS						PRODUCT MOVEMENT IN TIME AND SPACE				RISK MANAGEMENT				WELFARE							
		Yield	Technology adoption	Area under cultivation	Technical efficiency	Crop diversification	Product quality	Market access	Agricultural profits	Awareness / knowledge on market demand and prices	% of crop sold	Cross border trade	Inputs availability	Produce procurement services	Infrastructure (roads and facilities)	Food waste management	Food safety management	Yield risk	Price or market risk	Contract risk	Climate change risk	Family income	Food security and food access	Nutritional security	Poverty reduction	Education and healthcare	Employment conditions and opportunities	Environmental conservation or pollution reduction	Savings and expenditure
INPUTS	Input distribution network and management	31	12	8	6	4	0	7	21	0	2	0	4	3	2	0	2	0	1	1	0	13	11	1	4	2	2	2	4
	Information sharing and knowledge management	23	14	1	3	4	2	2	14	1	3	0	2	2	1	0	1	1	2	1	0	16	8	1	9	1	2	4	4
	Institutional credit and subsidies	41	15	17	8	6	1	4	26	2	2	0	4	2	0	0	1	3	5	2	0	24	11	2	10	4	1	2	9
	Land tenure and land improvement	11	4	4	3	5	0	1	4	0	1	0	1	2	0	0	0	1	0	0	0	3	6	2	2	0	0	2	0
	Water management	1	2	1	3	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
PRODUCTION, HARVEST, POST-HARVEST	Crop management services	8	5	1	5	2	0	2	4	0	1	1	1	1	0	0	0	0	0	1	1	9	3	0	3	1	2	1	3
	Dairy, meat and livestock related services	3	2	0	0	1	0	1	3	0	0	1	2	2	1	0	3	0	0	0	0	1	3	1	0	0	1	1	2
TRANSPORT AND STORAGE	Transport capacity, modernized storage, and training for good storage	2	2	0	1	2	0	2	5	0	0	0	1	1	2	0	1	0	0	0	0	3	3	0	1	1	1	0	2
POCESSING	Value addition services	5	1	0	2	1	1	1	1	0	0	1	1	1	1	0	1	0	0	0	0	6	1	0	1	0	1	1	0
	Product traceability and food safety services	0	0	0	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
MARKETING	Contract farming and industry tie- up	16	8	0	5	2	1	5	9	0	2	0	3	3	1	0	2	0	3	3	0	9	6	0	2	1	0	0	2
	Cerification along value chain	2	0	0	0	0	0	2	5	1	0	0	0	0	1	0	0	0	2	1	0	5	1	0	0	2	2	1	0
	Commodity markets and trading plateforms	4	3	1	1	0	0	5	4	0	0	0	0	0	0	0	0	0	0	1	0	1	2	0	3	1	3	0	2
	Market information networks and channels	3	1	0	1	1	0	2	4	1	1	0	1	1	0	0	0	0	1	0	0	3	3	1	1	0	0	0	0
GOVERNANCE AND INCLUSION	Interventions on gender, poverty, and social issues	0	0	1	0	1	0	0	2	0	0	0	0	0	1	0	0	0	0	0	0	1	0	1	0	0	0	0	1
	Value chain coordination	1	0	0	0	0	0	0	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0
	Trade policies	1	0	0	0	0	1	3	8	1	1	4	2	1	0	0	1	0	1	1	0	2	1	1	1	0	0	0	0

Note: Heatmaps by row are designed to highlight the frequency of occurrence by category (where dark green represents the largest frequency along the row and the red represents the lowest).

This heatmap shows the largest number of outcomes for each specific intervention category. For example, the second intervention category "Information sharing and knowledge management" resulted in change in the "Yield" outcome category in 23 studies. Being the highest number in that row, means that for this intervention change in yields resulted more than in any other outcome in the table.

Figure 21. Number of Studies Heatmap by Rows

Results by outcome categories: Looking along the columns of the EGM one can observe what outcomes may be more likely achieved by certain intervention category (Figure 22). Institutional credit and subsidies were the top intervention category impacting the largest number of outcomes, showing up in dark green in 16 out of 28 columns. More specifically, institutional credit and subsidies (the third row in the Figure 22) impact yield, technology adoption, area under cultivation, technical efficiency, crop diversification, agricultural profits, family income, food security, poverty reduction, and several other outcomes, more than any other intervention. All these categories are highlighted in dark green along the row corresponding to institutional credit and subsidies in Figure 22. Interventions in water management, value addition services, product traceability services, certification along the value chain, market information networks and channels, as well as interventions related to gender, poverty, social issues, and to value chain coordination were not among top categories impacting any of the 28 outcome categories.

		PRODUCTION OUTCOMES						MARKET REFORMS						PRODUCT MOVEMENT IN TIME AND SPACE				RISK MANAGEMENT				WELFARE							
		Yield	Technology adoption	Area under cultivation	Technical efficiency	Crop diversification	Product quality	Market access	Agricultural profits	Awareness / knowledge on market demand and prices	% of crop sold	Cross border trade	Inputs availability	Produce procurement services	Infrastructure (roads and facilities)	Food waste management	Food safety management	Yield risk	Price or market risk	Contract risk	Climate change risk	Family income	Food security and food access	Nutritional security	Poverty reduction	Education and healthcare	Employment conditions and opportunities	Environmental conservation or pollution reduction	Savings and expenditure
INPUTS	Input distribution network and management	31	12	8	6	4	0	7	21	0	2	0	4	3	2	0	2	0	1	1	0	13	11	1	4	2	2	2	4
	Information sharing and knowledge management	23	14	1	3	4	2	2	14	1	3	0	2	2	1	0	1	1	2	1	0	16	8	1	9	1	2	4	4
	Institutional credit and subsidies	41	15	17	8	6	1	4	26	2	2	0	4	2	0	0	1	3	5	2	0	24	11	2	10	4	1	2	9
	Land tenure and land improvement	11	4	4	3	5	0	1	4	0	1	0	1	2	0	0	0	1	0	0	0	3	6	2	2	0	0	2	0
	Water management	1	2	1	3	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
PRODUCTION, HARVEST, POST-HARVEST	Crop management services	8	5	1	5	2	0	2	4	0	1	1	1	1	0	0	0	0	0	1	1	9	3	0	3	1	2	1	3
	Dairy, meat and livestock related services	3	2	0	0	1	0	1	3	0	0	1	2	2	1	0	3	0	0	0	0	1	3	1	0	0	1	1	2
TRANSPO RT AND STORAGE	Transport capacity, modernized storage, and training for good storage	2	2	0	1	2	0	2	5	0	0	0	1	1	2	0	1	0	0	0	0	3	3	0	1	1	1	0	2
POCESSING	Value addition services	5	1	0	2	1	1	1	1	0	0	1	1	1	1	0	1	0	0	0	0	6	1	0	1	0	1	1	0
	Product traceability and food safety services	0	0	0	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
MARKETING	Contract farming and industry tie- up	16	8	0	5	2	1	5	9	0	2	0	3	3	1	0	2	0	3	3	0	9	6	0	2	1	0	0	2
	Cerification along value chain	2	0	0	0	0	0	2	5	1	0	0	0	0	1	0	0	0	2	1	0	5	1	0	0	2	2	1	0
	Commodity markets and trading plateforms	4	3	1	1	0	0	5	4	0	0	0	0	0	0	0	0	0	0	1	0	1	2	0	3	1	3	0	2
	Market information networks and channels	3	1	0	1	1	0	2	4	1	1	0	1	1	0	0	0	0	1	0	0	3	3	1	1	0	0	0	0
GOVERNANCE AND INCLUSION	Interventions on gender, poverty, and social issues	0	0	1	0	1	0	0	2	0	0	0	0	0	1	0	0	0	0	0	0	1	0	1	0	0	0	0	1
	Value chain coordination	1	0	0	0	0	0	0	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0
	Trade policies	1	0	0	0	0	1	3	8	1	1	4	2	1	0	0	1	0	1	1	0	2	1	1	1	0	0	0	0

Note: Heatmaps by column are designed to highlight the frequency of occurrence by category (where dark green represents the highest frequency along the column and the red represents the lowest).

This heatmap shows what category of intervention corresponds to the highest number of studies for each particular outcome. For example, second intervention category called "Information sharing and knowledge management" resulted in change in outcome called "Yield" in 23 studies. This is the third largest number along that column. The highest number is 41, which means that increase in "Yield" was achieved more frequently under intervention called "Institutional credit and subsidies".

Figure 22. Number of Studies Heatmap by Column

Looking along the columns of the EGM we should be able to identify interventions that are having higher impact on certain outcome category.

Some of the interventions are implemented in the marketing stage of the value chain and are to improve outcomes in the market expansion category, however selected studies do not report a lot of market related outcomes. This is indicated by dominating red color cells in the intersection of interventions of “Marketing” and resulting outcomes under “Market reforms” in Figure 22. Global value chains are becoming more buyer driven (Ponte and Gibbon, 2005), implying certifications and improved product quality should lead to improved product demand and ease of market access. However, Oya et al. (2017) argue that the impact of certification on producers’ welfare may not be significant. Current research indicates that interventions focused on input supply stage (distribution network and management and information sharing and knowledge management) have been more successful in achieving outcomes related to market access, than marketing stage interventions. To explain this, one should keep in mind that smallholder farmers, who make up most agricultural producers in developing countries, typically have plot sizes between 1.5 to 3 hectares (Sheahan and Barrett, 2017). Due to their limited output, these farmers may not be able to enter the market as individual sellers or may not be willing to bear the costs of certification. However, they may be likely to collaborate with processors to supply the required quality of produce. Additionally, while government and international organizations tend to support interventions in the input supply stage, such as subsidies, seed provision, and extension services, certification is typically handled by stakeholders themselves. This could result in a larger number of publications examining the input stage compared to the other stages of the value chain, because donor organizations and governments require publicity for the implemented projects.

Conclusions and Discussion

This study reveals gaps in agricultural value chain policies and identifies collaboration models that can support policy change. Specific questions addressed are:

- What are the extent and characteristics of existing empirical evidence regarding the policies that enhance AVC and improve enabling environments and what are the evidence gaps in the literature?
- What are the extent and characteristics of existing empirical evidence regarding the effectiveness of models of collaboration among local and international actors, including donors, private sector partners, academic institutions, and NGOs to support policy change?

Evidence gap map method was used to answer the first question. Second question was answered through thematic synthesis where text is coded to identify main themes emerging from the literature and frequencies of each theme are determined to come up with key findings.

The largest number of studies discuss interventions implemented in the input supply stage of the value chain; whereas the smallest number of studies addressed governance along value chains and interventions targeted toward gender, poverty, and social issues. These findings may imply there is a need for increased studies in the underexplored areas. Specifically, interventions in gender, poverty, and social issues may have high impact, because in low, middle, and upper-middle income countries the poverty rates are high and there is a large female involvement in agricultural activities. Adegbite and Machete (2020) mention that almost half of the agricultural workforce consists of female smallholders.

Implications for Researchers and Policymakers

Interventions resulted in the largest number of outcomes in production related outcomes' group, specifically in the yield category and in agricultural profit and family income. The group of outcomes related to risk management and food waste had the smallest frequency of occurrence, providing opportunities for future research.

If policymakers are constrained to certain types of interventions, it is helpful to understand what are the most likely outcomes that can be achieved through those interventions. For example, institutional credit and subsidies are influential interventions, because they impact more outcome categories than any other intervention. There is a gap in interventions related to product traceability services due to being mentioned in only two studies. Overall, among interventions related to the marketing stage of the value chain, contract-farming sub-category has been studied more than the other three sub-categories. The reason may be the nature of smallholder farming and their dependency on processors or retailers.

When it comes to the models of collaboration, the number of participants is far broader than just a farmer and a processor. Collaborating parties included farmers (or farmer organizations), industry partners (traders, processors), local research institutions, financial sector, governments (local or international), as well as development organizations. Most studies had positive outcomes from the collaboration implying collaboration can lead to successful interventions. Although in some cases collaboration was driven by the government or development organizations, the main driver for effective collaboration models was anticipated profits. The largest share of collaboration models was aimed at connecting farmers to processors (or traders) or at improving their ability to sell the product, which highlights the importance of access to market for smallholders. Although input supply stage of the value chain has the highest number of interventions, the highest percentage of studies discussing collaboration models was aimed at improving markets stage, followed by the production and post-harvest stage of the value chain. This suggests that when faced with limited resources, more attention should be directed towards improving market access, production, and post-harvest handling, rather than solely focusing on the input supply stage which may already be well-explored. The reason input supply stage interventions receive more attention could be due to the funding nature of interventions. For example, activities such as subsidies for agricultural loans, fertilizers, or new seed varieties are usually funded by governments and donor organizations, who encourage publicizing the intervention. However, interventions in marketing stage (including contract farming, certification, or establishment of a trading platform) are implemented primarily in pursuit of profit by collaborating parties who may not prioritize making the results of their collaboration available. It is also worth noting that the degree of attention and emphasis on the various stages of the value chain can differ depending on the specific context and challenges confronted by different agricultural systems and supply chains.

This research has limitations related to the scope of data collection. Evidence is collected only from studies published in English language and published after 2000. Moreover, study search was limited to several databases and websites, which implies there could be relevant studies that were not identified by the search. Researcher bias, although moderated by double-blind screening, could affect study selection results. Another limitation is related to method used. EGM is based only on evidence that is made available to the community. Some relevant evidence that was not published is not be included in the EGM. In addition,

CONCLUSIONS AND LIMITATIONS

This dissertation focuses on examining policies to improve education and knowledge to further development in the agricultural sector. Developing agricultural policies to supply food and fiber to the growing world population is crucial, regardless of whether they are implemented in educational institutions or in the field.

The overall objective of the dissertation is enhancing agricultural value chains through revealing factors affecting quality of agricultural education and highlighting gaps in policies that if addressed may increase the impact of interventions at different stages of agricultural value chains. To achieve this objective, three essays that address education and the current state of knowledge are conducted. The first two essays discuss the notion of grade increases versus inflation by examining factors that may impact grades, where the third essay identifies policies in agriculture that may impact agricultural productivity, profitability, market access, and farmers' welfare.

Policies in Agricultural Education

Agricultural education is crucial in transferring knowledge and skills to young people interested in pursuing careers in agriculture. In this regard, grading policies play a significant role in ensuring that students who complete agricultural courses meet certain standards. Grading policies are designed to evaluate students' knowledge and preparedness for the real-world challenges, which is why it is important that grades accurately signal the quality of education. In the presence of grade inflation comparison of students' grade point averages (GPAs) as a measure of their preparedness becomes inaccurate. Based on the data collected from College of Agriculture and Life Sciences (COALS) from 1985 to 2019, the first essay finds evidence of systematic increases in mean and standard deviation in class GPAs after controlling for the influence of institution, instructors, and student factors. In this essay, grades from classes in COALS are analyzed using hierarchical mixed effect model, which accounts for instructor and department specific differences.

Assuming GPAs are good measures of learning, the statistical significance of control variables lead to some recommendations to improved learning. Given both this study and previous studies findings (Skinner, 1985; Wheaton, Chapman, and Croft, 2016), administration may choose to schedule their degree-specific classes in the afternoon to increase student learning. In addition, to enhance learning outcomes, it is recommended to decrease the average class size (keeping in mind budgetary considerations), as smaller classes tend to yield higher GPAs. This is consistent with previous studies (Kokkelenberg, Dillon, and Christy, 2008). The gender of the instructor did not significantly influence grades. This finding is consistent with some studies (Figlio, Schapiro, and Soter, 2015; Ronco and Cahill, 2004), but deviates from others (Moore and Trahan, 1998; Kezim, Pariseau, and Quinn, 2005). Non-tenured track instructors and graduate students tend to grade higher. If non-tenured track instructors can devote more time to preparing for the class, then grades earned in their classes should reflect better learning (Figlio, Schapiro, and Soter 2015). However, if instructor inflated grades in the hopes of better student evaluations, which may be necessary to retain employment as many non-tenure track instructors are hired on temporary or part-time basis, then the increase in grades is attributed to grade inflation. If hiring non-tenure track instructors leads to grade inflation (Kezim, Pariseau, and Quinn, 2005; Sonner, 2000) such practices may be a worrisome policy to address decreasing budgets. Understanding the reasons behind increases in grades granted by non-tenured track instructors may be crucial for faculty recruitment policies.

Nearly all variables among students' characteristics are significant drivers of grades. A larger percentage of female students in class and larger percentages of students with high SAT scores and school rank increase class GPA. This suggests student recruitment policies (setting higher grade requirement for transfer students or increasing SAT threshold) may result in better prepared graduates. The drawback of this approach is rising admission standards could divert mature students, who do not meet these requirements, even though they may have the potential to excel academically (Griffin and Smithers, 1984).

Results of the hierarchical mixed effect model in the first essay revealed that different departments significantly (p-value < 0.05) contribute to the amount of variation in average GPA. This suggests it is necessary to investigate each department individually, to make better policy decisions. The second essay estimates individual models for 12 departments. In addition, data is split into two time periods (1989-2003 and 2004-2019) to compare the performance and significant factors within the same department across time. In the first period only four departments showed grade inflation, whereas in the second period this number goes to eight departments. Even among those that had grade inflation in both periods, there are substantial differences in significance of explanatory factors. For example, grades in Department of Soil and Crop Sciences (SCSC) and Department of Biological and Agricultural Engineering (BAEN) increased in both periods, however BAEN appears to have attracted students with better academic records (high school rank) where the ranking of SCSC students decreased. Similarly, Department of Agricultural Leadership, Education and Communication (ALEC) started with just about 300 students taking classes from this department in 1985, however increasing GPA to almost 3.8 in 90s, may have helped attract over 3000 students in classes by 2005. This suggests the grades might be used as an effective tool for recruiting students to take classes from a particular department.

Policies in Agricultural Value Chains

Agricultural value chains (AVCs) consist of multiple components: agricultural inputs supply, production, harvest, and post-harvest, transportation and storage, processing, and marketing stages. All stages require qualified labor force to ensure smooth operations and to improve governance along the AVC. Thorough understanding of interventions and resulting outcomes in each stage of the value chain supports informed decision making about future areas of intervention. The objectives of the third essay are collecting evidence on interventions in AVC and highlighting evidence gaps in the literature, as well as revealing models of stakeholder collaboration aimed at improving policy interventions in all stages of the AVCs.

Evidence Gap Map (EGM) method is used to systematically collect and analyze data to answer the first objective. EGM visually presents information that is known or not known on a specific topic; therefore, allowing organizations and agencies to target gaps potentially leading to higher impact policy interventions. The second objective is addressed via thematic synthesis of main themes emerging from the literature. Specifically, models of stakeholders' collaboration are identified and then grouped based on different characteristics.

One hundred ninety-three studies meet the criteria to be included in the EGM. Most frequently occurring interventions are related to input supply stage. The least frequently occurring interventions are related to governance along AVC and interventions targeted toward gender, poverty, and social issues. These finding implies there may be a need for increased interventions related to governance and gender, poverty and social issues, especially since there is a large female involvement in agricultural activities in the target countries. Based on the results of the study, if the policy goal is to increase yield, more likely interventions should be those related to the input supply stage. On the other hand, if policymakers prioritize increase in farmers' income and profits, they should implement policies aimed at promotion of high value crops cultivation, certification of produce and ensure better access to market.

When policymakers are constrained to certain types of interventions, it is essential to understand what are the most likely outcomes that can be achieved through those interventions. For example, institutional credit and subsidies are influential interventions, because they impact more outcome categories than any other intervention.

Among selected studies, the input supply stage of the value chain has the highest number of interventions, which implies that policy makers and researchers prioritize this stage of the value chain. However, stakeholder collaboration models were mainly centered around improved market access (contract farming, certification, or establishment of a trading platform). Collaborating parties were mainly farmers and the industry partners, who engaged into collaboration model primarily to increase profits. This suggests that when faced with limited resources, processors and farmers (as opposed to policy makers and researchers) may prefer improved market access, production, and post-harvest handling practices, than enhance activities related to input supply.

Limitations and Suggestions for Future Research

Data used for the grade inflation studies is from only one college in one university. Questions arise if the data is representative of other universities. However, the results may be generalized to other land grant universities with similar characteristics because they support findings of earlier studies that did not explicitly focus on colleges of agriculture. Study limitations encompass selection of variables included in the model. Some institutional variables (that could potentially contribute to student learning) such as modern equipment availability in the classroom or lab were not possible to obtain since some classrooms and even buildings were completely modified or transformed. Another limitation was individual student data is protected by the university, thus the class averages are used for all student-related variables.

Further research on grading patterns directly before and after instructors' promotion would increase understanding of the role of instructors in class GPAs. Such studies would provide insights on how the promotion process impacts grades, an important aspect of academic life. Another relevant research is conducting experimental studies on reciprocity in grading and student evaluations. The second period (2004-2019) analyzed in the second essay corresponds roughly to the years when baby boomers started teaching and millennials started attending college. One potential reason for increased grades in this period could be a push from students who belong to the millennial cohort that have a strong internal need to succeed and external pressure from parents to excel in competition. Examining different generation cohorts may be fruitful area of research. One additional aspect is investigating how the grading environment was impacted by the COVID-19 pandemic and to what extent these changes have persisted.

It is important to mention that the results of the EGM study are time and language bound. Information is drawn only from the evidence published in English language and after 2000. In addition, search criteria only included keyword and manual search in limited number of digital libraries and organizations websites, which implies there may be relevant studies that were not included. Snowballing or other search options were not used to expand the potential number of studies' pool. Another limitation is the necessity of making a wide range of assumptions, while evaluating impacts of interventions. For example, when evaluating extension advice, studies that report evidence assume the implementation went as it was planned, or farmers exhibited trust to the extension agent and exactly followed the advice. Impact evaluations usually do not include information on credibility of extension agent or trust farmers had towards the information provided by the extension agent. Finally, although systematic reviews were screened for quality, the internal or external validity of impact evaluations was not assessed.

Future research may target the gaps on the EGM, as these areas are underexplored in the literature. The outcome category associated with risk reduction had the least number of occurrences, and there were two specific subcategories, namely exchange rate risk and reduction in food waste, which were not observed. On the contrary, yield, agricultural profits, and family income were very frequently tested outcomes. This suggests agricultural policies and programs in low, middle, and upper-middle income countries tend to prioritize reducing poverty or increasing food production for the population rather than focusing on risk management objectives. There were very few studies addressing issues related to gender, poverty, and social issues, studies should consider focusing on those topics. In addition, studies show that value chain participants prioritize activities improving market access, production, and post-harvest handling, future studies could attend to possible improvements of those stages of the value chain.

REFERENCES

Aarhus University Library. 2022. Systematic Literature Searches - What is a Systematic Literature Search? *Aarhus University*, https://library.au.dk/en/researchers/systematicliteraturesearches. (Accessed April 25, 2023).

Abdelhedi, I. and S. Zouari. 2020. Agriculture and Food Security in North Africa: A Theoretical and Empirical Approach. *Journal of the Knowledge Economy* 11(1): 193-210.

Achen, A. and P. Courant. 2009. "What Are Grades Made Of?" *Journal of Economic Perspectives* 23(3): 77–92.

Adams, S. 2015. Do Employers Care About College Grades. *Forbes.* Internet site: https://www.forbes.com/sites/susanadams/2015/07/08/do-employers-care-about-college-grades/?sh=586e2d3a4b07. (Accessed November 9, 2021).

Adegbite, O. and C. Machete. 2020. Bridging the Financial inclusion gender gap in smallholder agriculture in Nigeria: An Untapped Potential for Sustainable Development. *World Development* 127:104755.

Adelman, C. 2004. Principal Indicators of Student Academic Histories in Postsecondary Education, 1972-2000. *Washington, DC: U.S. Department of Education, Institute of Education Sciences.* Internet site: https://eric.ed.gov/?id=ED483154. (Accessed September 7, 2018).

Adelman, C. 2008. Undergraduate Grades: A More Complex Story than Inflation. *Grade Inflation: Academic Standards in Higher Education* L. H. Hunt ed. Albany: State University of New York Press.

Archibald, R. and D. Feldman. 2012. The Anatomy of College Tuition. *The American Council on Education.* http://www.ntrp16.org/sites/default/files/Resources/ACE-%20The%20Anatomy%20of%20College%20Tuition.pdf. (Accessed June 24, 2022).

Aye, S. 2017. Rice Contract Farming System and its Impact on Smallholder Livelihood in Myanmar. *Margraf Verlag*. ISBN: 9783823617402.

Bachan, R. 2018. The Drivers of Degree Classification. Research Report, *University of Brighton*. Available at https://research.brighton.ac.uk/en/publications/the-drivers-of-degree-classification. (Accessed September. 21, 2021).

Baird, A., J. Carter, and J. Roos. 2019. Seeking Evidence of Grade Inflation at For-profit Colleges and Universities. *Sociological Focus* 52(4): 343-358.

Bar, T., V. Kadiyali, and A. Zussman. 2009. Grade Information and Grade Inflation: The Cornell Experiment. *The Journal of Economic Perspectives* 23(3): 93-108.

Barker, S. and N. Pomerantz. 2000. Impact of Learning Communities on Retention at a Metropolitan University. *Journal of College Student Retention: Research, Theory & Practice* 2(2): 115-126.

Becker, H., B. Geer, and E. Hughes. 1968. Making the Grade: The Academic Side of College life. *New York: John Wiley.* Ebook published 2017 available at https://www.taylorfrancis.com/books/mono/10.4324/9780203786635/making-grade-howard-becker. (Accessed September 9, 2022).

Beenstock, M. and D. Feldman. 2016. Decomposing University Grades: A Longitudinal Study of Students and Their Instructors. *Studies in Higher Education* 43(1): 114-133.

Bejar, I. and E. Blew. 1981. Grade Inflation and Validity of the Scholastic Aptitude Test. *American Educational Research Journal* 18(2): 143-156.

Bergtold, J., E. Yeager, and T. Griffin. 2016. Academic, Demographic and Spatial Factors in the Classroom Affecting Student Performance in Principles of Agricultural Economics Courses. *Paper presented at the AAEA Annual Meeting, Boston, MA*. Internet site: https://econpapers.repec.org/paper/agsaaea16/235930.htm. (Accessed January 21, 2022).

Bond, T. and K. Mumford. 2019. The Causes and Consequences of Purdue Grade Inflation. *Krannet School of Management Research Center in Economics*. https://www.purdue.edu/senate/documents/meetings/Causes-and-Consequences-of-Purdue-Grade-Inflation.pdf. (Accessed November 7, 2021).

Boretz, E. 2004. Grade Inflation and the Myth of Student Consumerism. *College Teaching* 52(2):42-46.

Brighouse, H. 2008. Grade Inflation and Grade Variation: What's All the Fuss About? In L. H. Hunt (Ed.), Grade Inflation: Academic Standards in Higher Education (pp. 73–91). New York Press. ISBN: 978-0-7914-7497-6

Butcher, K., P. McEwan, and A. Weerapana. 2014. The Effects of an Anti-Grade Inflation Policy at Wellesley College. Journal of Economic Perspectives 28 (3): 189–204.

Cattaneo, M., P. Maloghetti, M. Meoli, and S. Paleari. 2016. University Spatial Competition for Students: The Italian Case. *Regional Studies* 51(5): 750-764.

Causey, J., A. Pevitz, M. Ryu, A. Scheetz, and D. Shapiro. 2022. Completing College: National and State Report on Six Year Completion Rates for Fall 2015 Beginning Cohort (Signature Report 20). *National Student Clearinghouse Research Center.* https://nscresearchcenter.org/wp-content/uploads/Completions_Report_2021.pdf. (Accessed December 26, 2022).

Chen, K., Z. Hansen, and S. Lowe. 2021.Why Do We Inflate Grades? The Effect of Adjunct Faculty Employment on Instructor Grading Standards. *Journal of Human Resources* 56: 878-921.

Chowdhury, F. 2018. Grade Inflation: Causes, Consequences and Cure. *Journal of Education and Learning* 7(6): 86-92.

Clayton, K. and W. Preston. 2003. The Political Economy of Eifferentiating Markets: Facing Reality Inside the US Department of Agriculture. *American Journal of Agricultural Economics* 85(3): 737-741.

College Board. 2015. SAT Total Group Profile Report. Internet site: https://secure-media.collegeboard.org/digitalServices/pdf/sat/total-group-2015.pdf. (Accessed October 16, 2022).

Covidence, 2022. Better Systematic Review Management. https://www.covidence.org. (Accessed through May 20, 2023).

Curran, T. and A. Hill. 2019. Perfectionism is Increasing Over Time: A Meta-analysis of Birth Cohort Differences from 1989 to 2016. *Psychological Bulletin,* 145(4): 410–429.

Cuseo, J. 2007. The Empirical Case Against Large Class Size: Adverse Effects on the Teaching. Learning, and Retention of First-Year Students. *Journal of Faculty Development* 21(1): 5-21.

Danilowicz-Gösele K., K. Lerche, J. Meya, and R. Schwager. 2017. Determinants of Students' Success at University. *Education Economics* 25(5): 513-532.

DATAUSA. 2020. Agriculture. Internet site: https://datausa.io/profile/cip/agriculture. (Accessed September 21, 2021).

Denning, J., E. Eide, K. Mumford, R. Patterson, and M. Warnick. 2022. Lower Bars, Higher College GPAs: How Grade Inflation is Boosting College Graduation Rates. *Education Next* 22(1): 56-62.

Diette, T. and M. Raghav. 2015. Class Size Matters: Heterogeneous Effects of Larger Classes on College Student Learning. *Eastern Economic Journal; Basingstoke* 41(2): 273-283

Diette, T. and M. Raghav. 2016. A Student's Dilemma: Is There a Trade-off Between a Higher Salary or Higher GPA. *Education Economics* 24(6):1-10.

Diez-Roux, A. 1998. Bringing Context Back into Epidemiology: Variables and Fallacies in Multilevel Analysis. *American Journal of Public Health* 88(2): 216–222.

Do, M. and S. Park. 2018. New Rural Development and Hierarchical Governance in Vietnam: Impacts of Government Support on Rural Households' Income Using a Hierarchical Linear Modelling. *Agris On-Line Papers in Economics & Informatics* 10(4).

Douglas, L. 2004. Multilevel Modeling (Quantitative Applications in the Social Sciences). *Second Multilevel Edition*. Thousand Oaks, CA: *Sage Publishing*. SBN-13. 978-0761928799.

Eiszler, C. 2002. College Student's Evaluations of Teaching and Grade Inflation. *Research in Higher Education* 43(4): 483-501.

Figlio, D., M. Schapiro, and K. Soter. 2015. Are Tenure Track Professors Better Teachers? *The Review of Economics and Statistics* 97(4): 715-724.

Flynn, J. 1987. Massive IQ Gains in 14 Nations: What IQ Tests Really Measure. *Psychological Bulletin*, 101(2):171–191.

Fossati, F., A. Wilson, and G. Bonoli. 2020. What Signals Do Employers Use When Hiring? Evidence from a Survey Experiment in the Apprenticeship Market. *European Sociological Review* 36(5): 60–779.

Freeman, D. 1999. Grade Divergence as a Market Outcome. *The Journal of Economic Education* 30(4): 344–351.

Gero, A. and A. Egbendewe. 2020. Macroeconomic Effects of Semi-subsistence Agricultural Productivity Growth: Evidence from Benin and Extension to the WAEMU Countries. *Scientific African* 7, e00222.

Glodstein, H. and N. Noboken. 2011. Multilevel Statistical Models. 4th edition. *John Wiley & Sons Publishing:* p. 11.

GovInfo. 2021. College Completion Fund Act of 2021. *United States Government Publishing Office (GPO) Information Services*. https://www.govinfo.gov/app/details/BILLS-117s2755is. (Assessed 12 June, 2022).

Greene, W. 2012. Econometric Analysis. 7th edition, *Pearson*, London. ISBN: 978-0131395381

Griffin, A. and A. Smithers. 1984. Prospective Mature Students: Those Who Did Not Apply. Studies in Higher Education 9(1): 73-81. DOI:10.1080/03075078412331378933

Haladyna, T., S. Nolen, and N. Haas. 1991. Raising Standardized Achievement Test Scores and the Origins of Test Score Pollution. *Educational Researcher* 20(5): 2-7.

Hanson, M. 2022. College Enrollment & Student Demographic Statistics. *Education Data Initiative.* https://educationdata.org/college-enrollment-statistics. (Accessed June 23, 2022).

Hartnett, R. and J. Centra. 1977. The Effects of Academic Departments on Student Learning. *The Journal of Higher Education* 48(5): 491-507.

Henebry, K. 2010. The Impact of Class Schedule on Student Performance in a Financial Management Course. *Journal of Education for Business* 73(2):114-120.

Hermanowicz, J. and D. Woodring. 2019. The Distribution of College Grades across Fields in the Contemporary University. *Innovative Higher Education* 44(6): 497–510.

Hernández-Julián, J. and A. Looney. 2016. Measuring Inflation in Grades: An Application of Price Indexing to Undergraduate Grades. *Economics of Education Review* 55, C220-232.

Herron, M. and Z. Markovich. 2017. Student Sorting and Implications for Grade Inflation. *Rationality and Society* 29(3):355–386.

Higgins J., J. Thomas, J. Chandler, M. Cumpston, T. Li, M. Page, and V. Welch (editors). 2022. Cochrane Handbook for Systematic Reviews of Interventions, version 6.3 (updated February 2022). *Cochrane, 2022.* www.training.cochrane.org/handbook. (Accessed May 22, 2022)

Hoffmann, F. and P. Oreopoulos. 2009. A Professor Like Me: The Influence of Instructor Gender on College Achievement. *Journal of Human Resources* 44(2): 479-494

Howard, G. and S. Maxwell. 1980. Correlation between Student Satisfaction and Grades: A Case of Mistaken Causation? *Journal of Educational Psychology* 72(6): 810–820.

Howe, W. and N. Strauss. 2000. Millennials rising: The Next Great Generation. New York, NY: Vintage Books. ISBN 13: 9780375707193.

Hoxby, C. 2000. The Effects of Geographic Integration and Increasing Competition in the Market for College Education. *NBER Working Paper*, 6323. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.195.6661&rep=rep1&type=pdf. (Accessed June 27, 2022).

Humphrey, J. and H. Schmitz. 2001. Governance in Global Value Chains. *IDS bulletin* 32(3): 19-29.

Jaschik, S. 2008. Students Fail and Professor Loses Job. *Inside Higher Ed*. Internet site: https://www.insidehighered.com/news/2008/05/14/students-fail-and-professor-loses-job. (Accessed September 13, 2018).

Jefferson, F., N. Gowar, and M. Naef. 2019. English Universities in Crisis: Markets without Competition. *Bristol University Press.* ISBN: 9781529202274.

Jephcote, C., E. Medland, and S. Lygo-Baker. 2021. Grade Inflation versus Grade Improvement: Are Our Students Getting More Intelligent? *Assessment & Evaluation in Higher Education*, 46(4): 547-571.

Jewell, R. and M. McPherson. 2012. Instructor-Specific Grade Inflation: Incentives, Gender, and Ethnicity. *Social Science Quarterly* 93(1): 95–109.

Jewell, R., M. McPherson, and M. Tieslau. 2013. Whose Fault is It? Assigning Blame for Grade Inflation in Higher Education. *Applied Economics* 45(9): 1185–1200.

Jones, N. and B. Rosenberg. 2018. Program Theory of Change. In *SAGE Encyclopedia of Educational Research, Measurement, and Evaluation* (pp. 1315-1318). *SAGE Publications*. ISBN: 9781506326153.

Joyce, T., S. Crockett, D. Jaeger, O. Altindag, and S. O'Connell. 2014. Does Classroom Time Matter? A Randomized Field Experiment of Hybrid and Traditional Lecture Formats in Economics. *NBER Working Paper* No. 20006. Internet site: http://www.nber.org/papers/w20006.ack. (Accessed September 10, 2018).

Juola, A. 1979. Grade Inflation in Higher Education: What Can or Should be Done? Paper presented at the *Annual Meeting of National Council on Measurement in Education, San Francisco, California*. Internet site: https://eric.ed.gov/?id=ED129917. (Accessed September 7, 2018).

Kapitanoff, S. and C. Pandey. 2017. Stereotype Threat, Anxiety, Instructor Gender, and Underperformance in Women. *Learning in Higher Education* 18(3): 213 –229.

Kamber, R. and M. Biggs. 2004. Grade Inflation: Metaphor and Reality. *Journal of Education* 184(1): 31- 37.

Kaplinsky, R. and M. Morris. 2000. A Handbook for Value Chain Research. Vol. 113. Brighton: University of Sussex, Institute of Development Studies.

Kezim, B., S. Pariseau, and F. Quinn. 2005. Is Grade Inflation Related to Faculty Status? *Journal of Education for Business* 80(6): 358-364.

Kokkelenberg, E., M. Dillon, and S. Christy. 2008. The Effects of Class Size on Student Grades at a Public University. *Economics of Education Review* 27(2): 221-223.

Kobrin, J., B. Patterson, E. Shaw, K. Mattern, and S. Barbuti. 2008. Validity of the SAT for Predicting First-Year College Grade Point Average. *Research Report No. 2008-5*. New York. The College Board. https://eric.ed.gov/?id=ED563202. (Accessed March 13, 2020).

Kohn, A. 2002. The Dangerous Myth of Grade Inflation. *The Chronicle of Higher Education* 49(11): B7-B9.

Kolevzon, M. 1981. Grade Inflation in Higher Education: A Comparative Study. *Research in Higher Education* 15(3): 195-212.

Kostal, J., N. Kuncel, and P. Sackett. 2016. Grade Inflation Marches on: Grade Increases from the 1990s to 2000s. *Educational Measurement: Issues and Practice* 35(1): 11-20.

Kuh, G. and S. Hu. 1999. Unraveling the Complexity of the Increase in College Grades from the Mid-1980s to the Mid-1990s. *Educational Evaluation and Policy Analysis* 21(3): 297-320.

Kuh, G., J. Kinzie, J. Buckley, B. Bridges, and J. Hayek. 2006. What Matters to Student Success: A Review of the Literature. *Commissioned Report for the National Symposium on Postsecondary Student Success.* https://nces.ed.gov/npec/pdf/kuh_team_report.pdf. (Accessed February 11, 2022).

Larsen, K., R. Kim, and F. Theus (Eds.). 2009. Agribusiness and Innovation Systems in Africa. *World Bank Institute*, 1818 H Street NW, Washington, DC 20433, USA. ISBN: 978-0-8213-7944-8.

Lim S. 2021. Global Agricultural Value Chains and Structural Transformation. *NBER Working Paper* No. w29194. Available at SSRN: https://ssrn.com/abstract=3913837. (Accessed February 10, 2022).

Liu, H., Z. Yao, L. Zeng, and J. Luan. 2019. An RFID and Sensor Technology-based Warehouse Center: Assessment of New Model on a Superstore in China. *Assembly Automation* 39(1): 86-100.

Marbouti, F., A. Shafaat, J. Ulas, and H. Diefes-Dux. 2018. Relationship Between Time of Class and Student Grades in an Active Learning Course. *Journal of Engineering Education* 107(3): 468-490.

Marini, J., E. Shaw, L. Young, and M. Ewing. 2018. Getting to Know your Criteria: Examining College Course Grades and GPAs over Time. *Research Report College Board.* Available at https://files.eric.ed.gov/fulltext/ED582569.pdf. (Accessed September 21, 2021).

Marsh H. and L. Roche. 2000. Effects of Grading Leniency and Low Workload on Students' Evaluations of Teaching: Popular Myth, Bias, Validity, or Innocent Bystanders? *Journal of Educational Psychology* 92(1): 202–228.

Mason, N., A. Wineman, and S. Tembo. 2020. Reducing Poverty by 'Ignoring the Experts'? Evidence on Input Subsidies in Zambia. *Food security* 12: 1157-1172.

McAllister, C., X. Jiang, and F. Aghazadeh. 2008. Analysis of Engineering Discipline Grade Trends. *Assessment and Evaluation in Higher Education* 33(2): 167–178.

McElroy, T. and L. Mosteller. 2006. The Influence of Circadian Type, Time of Day and Class Difficulty on Students' Grades. *Electronic Journal of Research in Educational Psychology* 4(3): 611-622.

McGowen, M. and G. Davis. 2022. A Numerical Indicator of Student Cognitive Engagement and Mathematical Growth. *International Electronic Journal of Mathematics Education* 17(1). Article No: em0669. https://doi.org/10.29333/iejme/11473. (Accessed December 11, 2022).

Merrow, J. 2004. Grade Inflation: It's Not Just an Issue for the Ivy League. *Carnegie Perspectives*. ERIC Number: ED498995. https://files.eric.ed.gov/fulltext/ED498995.pdf. (Accessed June 2, 2022).

Minaya, V. 2020. Do Differential Grading Standards Across Fields Matter for Major Choice? Evidence from a Policy Change in Florida. *Research in Higher Education* 61(8):943-965.

Moher, D., A. Liberati, J. Tetzlaff, and D. Altman. 2009. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: the PRISMA Statement." Annals of Internal Medicine 151(4): 264-269.

Moore, M. and R. Trahan. 1998. Tenure Status and Grading Practices. *Pacific Sociological Association* 41(4): 775- 781.

Mostrom, A. and P. Blumberg. 2012. Does Learning-Centered Teaching Promote Grade Improvement? *Innovative Higher Education* 37(5): 397-405.

Muflikh Y., C. Smith, and A. Aziz. 2021. A Systematic Review of the Contribution of System Dynamics to Value Chain Analysis in Agricultural Development. *Agricultural Systems*, Volume 189, 103044, ISSN 0308-521X, https://doi.org/10.1016/j.agsy.2020.103044. (Accessed September 27, 2022).

Muyombano, E. and Espling, M. 2020. Land Use Consolidation in Rwanda: The Experiences of Small-Scale Farmers in Musanze District, Northern Province. *Land Use Policy*, 99, 105060.

Norton, R. 2004. Agricultural Development Policy. *Wiley*, 1st edition, ISBN-13: 978-0470857793: 8-9.

Norton, R. 2016. The Competitiveness of Tropical Agriculture: A Guide to Competitive Potential with Case Studies. *Academic Press*. ISBN 978-0-12-805312-6.

Nye, B., L. Hedges, and S. Konstantopoulos. 2001. Are Effects of Small Classes Cumulative? Evidence From a Tennessee Experiment. *The Journal of Educational Research* 94(6): 336-345.

O'Connor, L. 2020. The Decline and Fall of Grade Deflation. *The Daily Princetonian.* Internet site: https://www.dailyprincetonian.com/article/2020/01/the-decline-and-fall-of-grade-deflation. (Accessed September 30, 2020).

O'Dea, R., M. Lagisz, M. Jennions, and S. Nakagawa. 2018. Gender Differences in Individual Variation in Academic Grades Fail to Fit Expected Patterns for STEM. *Nature Communications* 9(1): 1-8.

Okry, F., P. Van Mele, E. Nuijten, P. Struik, and R. Mongbo. 2011. Organizational Analysis of the Seed Sector of Rice in Guinea: Stakeholders, Perception and Institutional Linkages. *Experimental Agriculture* 47(1):137-157.

Oelofse, S. and A. Nahman. 2013. Estimating the Magnitude of Food Waste Generated in South Africa. W*aste Management & Research* 31(1): 80-86.

Opstad, L. 2020. Why are There Different Grading Practices Based on Students' Choice of Business Major? *Educational Process International Journal* 9(1): 43-57.

Ost, B. 2010. The Role of Peers and Grades in Determining Major Persistence in Sciences. *Economics of Education Review* 29(6): 923−934.

Oya, C., F. Schaefer, D. Skalidou, C. McCosker, and L. Langer. 2017. Effects of Certification Schemes for Agricultural Production on Socio-Economic Outcomes in Low-and Middle-Income Countries: A Systematic Review. *Campbell Systematic Reviews* 13(1): 1–346.

Park, S. and E. Lake. 2005. Multilevel Modeling of a Clustered Continuous Outcome: Nurses' Work Hours and Burnout. *Nursing Research* 54(6): 406–413.

Pattison, E., E. Grodsky, and C. Muller. 2013. Is the Sky Falling? Grade Inflation and the Signaling Power of Grades. *Educational Researcher* 42(5): 259–265.

Pavez, I., J. Codron, P. Lubello, and M. Florêncio. 2019. Biosecurity Institutions and the Choice of Contracts in International Fruit Supply Chains. *Agricultural Systems* 176, 102668.

Peace II, J. 2017. How Employers Can Stanch the Hemorrhaging of Collegiate GPA Credibility. *Business Horizons* 60(1): 35-43.

Ponte, S. and P. Gibbon. 2005. Quality Standards, Conventions and the Governance of Global Value Chains. *Economy and Society,* 34(1): 1-31.

Popov, S. and D. Bernhardt. 2013. University Competition, Grading Standards, and Grade Inflation. *Economic Inquiry* 51(3): 1764-1778.

Protopop, I. and A. Shanoyan. 2016. Big Data and Smallholder Farmers: Big Data Applications in the Agri-food Supply Chain in Development Countries. *International Food and Agribusiness Management* Association 19(A): 1-18.

Rashid, S., N. Tefera, N. Minot, and G. Ayele. 2013. Can Modern Input Use be Promoted without Subsidies? An Analysis of Fertilizer in Ethiopia. *Agricultural Economics* 44(6): 595-611.

Rask, K. 2010. Attrition in STEM Fields at a Liberal Arts College: The Importance of Grades and Pre-collegiate Preferences. *Economics of Education Review* 29(6): 892 – 900.

Reames, E. and C. Bradshaw. 2009. Block Scheduling Effectiveness: A 10-Year Longitudinal Study of One Georgia School System's Test Score Indicators. *Georgia Educational Researcher* 7(1): article 2.

Rezaei M. and B. Liu. 2017. Food Loss and Waste in the Food Supply Chain. *International Nut and Dried Fruit Council, Reus, Spain,* pp. 26-27. Available online: https://www.fao.org/3/bt300e/BT300E.pdf (Accessed February 25, 2023).

Richardson, W., M. Wilson, J. Nishikawa, and R. Hayward. 1995. The Well-Built Clinical Question: A Key to Evidence-Based Decisions. *ACP Journal Club* 123: A12-13.

Rojstaczer, S. and C. Healy. 2010. Grading in American Colleges and Universities. Internet site: http://gradeinflation.com/tcr2010grading.pdf. (Accessed September 7, 2018).

Rojstaczer, S. and C. Healy. 2012. Where A is Ordinary: The Evolution of American College and University Grading, 1940 – 2009. *Teachers College Record* 114(7): 1-23.

Ronco, S. and J. Cahill. 2004. Does it Matter Who's in the Classroom? Effect of Instructor Type on Student Retention, Achievement and Satisfaction. Paper presented at the *44th Annual Forum of the Association for Institutional Research, Boston, Massachusetts*. Internet site: https://eric.ed.gov/?id=ED491000. (Accessed September 10, 2018).

Rothstein, J. 2004. College Performance Predictions and the SAT. *Journal of Econometrics* 121(1-2): 297-317.

Sabot, R. and J. Wakeman-Linn. 1991. Grade Inflation and Course Choice. *The Journal of Economic Perspectives* 5(1): 159–170.

Schaeffer, C. 2022. 10 Facts about Today's College Graduates. *Pew Research Center*. https://www.pewresearch.org/fact-tank/2022/04/12/10-facts-about-todays-college-graduates/ (Accessed June 24, 2022).

Schutz, K., B. Drake, J. Lessner, and G. Hughes. 2015. A Comparison of Community College Full-Time and Adjunct Faculties' Perceptions of Factors Associated with Grade Inflation. *The Journal of Continuing Higher Education* 63(3): 180-192.

Sheahan, M. and C. Barrett. 2017. Ten Striking Facts about Agricultural Input Use in Sub-Saharan Africa. *Food Policy* 67: 12-25.

Shinn, M. and B. Rapkin. 2000. *Cross-level Research without Cross-ups in Community Psychology*. Handbook of Community Psychology, J. Rappaport and E. Seidman, eds. Boston, MA.: *Kluwer Academic Publishers*.

Shobana, V. and S. Gandhimathi. 2015. Indirect Agricultural Advances in Pre and Post Reform Period. *International Journal of Social Science and Interdisciplinary Research* 4(6): 56-62.

Skinner, N. 1985. University Grades and Time of Day of Instruction. *Bulletin of the Psychonomic Society* 23(1):67.

Smyth, S., P. Phillips, and W. Kerr. 2016. EU Failing FAO Challenge to Improve Global Food Security. *Trends in Biotechnology* 34(7): 521-523.

Snilstveit, B., R. Bhatia, K. Rankin, and B. Leach. 2017. A Starting Point for Strategic Evidence Production and Use. *International Initiative for Impact Evaluation (3ie) Evidence Gap Maps*. Working Paper 28. https://www.3ieimpact.org/evidence-hub/publications/working-papers/3ie-evidence-gap-maps-starting-point-strategic-evidence (Accessed June 28, 2022).

Solanki, S. and D. Xu. 2018. Looking Beyond Academic Performance: The Influence of Instructor Gender on Student Motivation in STEM Fields. *American Educational Research Journal* 55(4): 801-835.

Sonner, B. 2000. 'A is for Adjunct': Examining Grade Inflation in Higher Education. *Journal of Education for Business* 76(1): 5-8.

StataCorp. 2017. Stata Multilevel Mixed-Effects Reference Manual. Release 15. *ISBN-10: 1-59718-244-3*. https://www.stata.com/manuals15/me.pdf. (Accessed March 8, 2020).

Stratton, R., S. Myers, and R. King. 1994. Faculty Behavior, Grades, and Student Evaluations. *Journal of Economic Education* 25(1): 5-15.

Suslow, S. 1976. A Report on an Interinstitutional Survey of Undergraduate Scholastic Grading 1960s to 1970s. Internet site: https://eric.ed.gov/?id=ED129187. (Accessed September 7, 2018).

Teixeira, P., V. Rocha, R. Biscaia, and M. Cardoso. 2014. Revenue Diversification in Public Higher Education: Comparing the University and Polytechnic Sectors. *Public Administration Review* 74(3): 398-412.

Tomkin, J. and M. West. 2022. STEM Courses are Harder: Evaluating Inter-course Grading Disparities with a Calibrated GPA Model. *International Journal of STEM Education* 9, article number 27. https://doi.org/10.1186/s40594-022-00343-1. (Accessed November 23, 2022).

Trahan, L., K. Stuebing, M. Hicock, and J. Fletcher. 2014. The Flynn Effect: A Meta-Analysis. *Psychology Bulletin* 14(5): 1332–1360.

Trienekens, J. 2011. Agricultural Value Chains in Developing Countries: A Framework for Analysis. *International Food and Agribusiness Management Review* 14(2): 51-83.

USDA. 2020. USDA Food for Progress Learning Agenda: Trade Expansion and Agricultural Market Development. https://www.fas.usda.gov/sites/default/files/2020-03/learning_agenda_final.pdf (Accessed September 14, 2022).

USDA Economic Research Service. 2017. Low-income Countries Had the Highest Rate of Agricultural Growth, but Middle-income Countries Had the Highest Rate of Productivity Growth. https://www.ers.usda.gov/data-products/chart-gallery/gallery/chart-detail/?chartId=85381 (Accessed February 22, 2023).

US Census Bureau. 2022. Educational Attainment in the United States: 2021. https://www.census.gov/data/tables/2021/demo/educational-attainment/cps-detailed-tables.html. (Accessed June 23, 2022).

Van Campenhout, B., B. Minten, and J. Swinnen. 2021. Leading the Way–Foreign Direct Investment and Dairy Value Chain Upgrading in Uganda. *Agricultural Economics* 52(4): 607-631.

Voyer, D. and S. Voyer. 2014. Gender Differences in Scholastic Achievement: A Meta-Analysis. *Psychological Bulletin* 140(4): 1174–1204.

Voice of America. 2019. US College Leaders See Too Much Competition Ahead. https://www.voanews.com/a/student-union_us-college-leaders-see-too-much-competition-ahead/6180386.htm. (Accessed April 10, 2022).

Wahlstrom, K., V. Hendrix, and J. Frederickson. 1998. School Start Time Study. Technical Report, Volume II: Analysis of Student Survey Data. Minneapolis, MN: *University of*

Minnesota, Center for Applied Research and Educational Improvement. Internet site: https://conservancy.umn.edu/bitstream/handle/11299/4249/SST-1998.Vol.II-Student%20Survey%20results%20%281%29.pdf?sequence=7&isAllowed=y. (Accessed October 17, 2020).

Wang, M. and J. Degol. 2016. School Climate: A Review of the Construct, Measurement, and Impact on Student Outcomes. *Educational Psychology Review* 28(2): 315-352.

Warr, P. and W. Suphannachart. 2021. Agricultural Productivity Growth and Poverty Reduction: Evidence from Thailand. *Journal of Agricultural Economics* 72(2): 525-546.

Watjatrakul, B. 2014. Factors Affecting Students' Intentions to Study at Universities Adopting the 'Student-as-Customer' Concept. *International Journal of Educational Management* 28(6): 676-693.

Weston, T., E. Seymour, A. Koch, and B. Drake. 2019. Weed-Out Classes and Their Consequences. In: Seymour, E., Hunter, AB. (eds) *Talking about Leaving Revisited:* 197–243. https://doi.org/10.1007/978-3-030-25304-2_7. (Accessed October 17, 2022).

Westrick, P., H. Le, S. Robbins, J. Radunzel, and F. Schmidt. 2015. College Performance and Retention: A Meta-analysis of the Predictive Validities of ACT Scores, High School Grades, and SES. *Educational Assessment* 20: 23-45.

Wheaton, A., D. Chapman, and J. Croft. 2016. School Start Times, Sleep, Behavioral, Health, and Academic Outcomes: A Review of the Literature. *Journal of School Health* 86(5): 363–381.

Weiss, C. 1995. Nothing as Practical as a Good Theory: Exploring Theory-Based Evaluation for Comprehensive Community Initiatives for Children and Families. In J. P. Connell, A. C. Kubisch, L. B. Schorr, & C. H. Weiss (Eds.), *New Approaches to Evaluating Community Initiatives: Concepts, Methods, and Contexts* (pp. 65-92). Washington, DC: Aspen Institute.

Witteveen, D. and P. Attewell. 2020. The STEM Grading Penalty: An Alternative to the "Leaky Pipeline" Hypothesis. *Science Education* 104(4): 714-735.

Wong, H., X. Wei, H. Kahsay, Z. Gebreegziabher, C. Gardebroek, D. Osgood, and R. Diro. 2020. Effects of Input Vouchers and Rainfall Insurance on Agricultural Production and Household Welfare: Experimental Evidence from Northern Ethiopia. *World Development* 135:105074.

World Bank, 2022. Agriculture and Food Overview. https://www.worldbank.org/en/topic/agriculture/overview (Accessed February 23, 2023).

World Bank Blogs, 2019. New Country Classifications by Income Level: 2018-2019. https://blogs.worldbank.org/opendata/new-country-classifications-income-level-2018-2019. (Accessed March 10, 2023).

Wu, L. 2019. Mixed Effect Models for Complex Data. *Chapman and Hall/CRC, London*. ISBN: 9780367384913: (pp 44-45).

Zadawa, A. and A. Omran. 2020. Rural Development in Africa: Challenges and Opportunities. In: Omran, A., Schwarz-Herion, O. (eds) Sustaining our Environment for Better Future. *Springer, Singapore.* https://doi.org/10.1007/978-981-13-7158-5_3. (Accessed October 23, 2022).

APPENDIX A - TABLES

Table A1. Variable Description and Summary Statistics for COALS Model

Variable Name	Variable Description	Mean	Standard Deviation	Minimum Value	Maximum Value
	Dependent Variable				
GPA	Unweighted class mean GPA	3.216	0.490	1.2	4
StDev	Standard deviation of the grades received in the class	0.774	0.303	0	1.952
	Grade Inflation Variable				
Ln trend	Natural logarithm of trend as given by semester	3.559	0.652	0	4.263
	Institutional Variables				
Department	Name of the department used as level (13 departments and a study program)				
Morning	Equals 1 if class starts before 12:01, 0 otherwise (dropped to avoid perfect collinearity)	0.623	0.485	0	1
Afternoon	Equals 1 if class starts between 12:01 to 15:59, 0 otherwise	0.356	0.479	0	1
Evening	Equals 1 if class starts at 16:00 or later, 0 otherwise	0.021	0.144	0	1
Meet 1	Equals 1 if the class meets once per week – usually class duration is 2.5 hours for a three-credit class, 0 otherwise	0.288	0.452	0	1
Meet 2	Equals 1 if the class meets twice per week – usually class duration is 75 minutes for a three-credit class, 0 otherwise	0.487	0.500	0	1
Meet 3	Equals 1 if the class meets three times per week – usually class duration is 50 minutes for a three-credit class, 0 otherwise (dropped to avoid perfect collinearity)	0.226	0.418	0	1

Lower division	Equals 1 if the class is listed as a 100 or 200 level class, 0 otherwise (dropped to avoid perfect collinearity)	0.232	0.422	0	1
Upper division	Equals 1 if the class is listed as a 300 or 400 level class, 0 otherwise	0.768	0.422	0	1
Total Students	Number of students receiving a grade A – F and no grades (see share below) in the class	51.084	50.334	5	349
Low_credit	Equals 1 if the class is 1 or 2 credit hours, 0 otherwise – very few classes are 2 credits	0.257	0.437	0	1
High_credit	Equals 1 if the class is 3 credit hours or more, 0 otherwise very few classes have more than3 credits	0.742	0.437	0	1
Semester	Equals 1 for classes held in the Fall and 0 for Spring classes	0.514	0.500	0	1
Y85-88	Equals 1 if the year is 1985, 1986 or 1987, 0 otherwise	0.033	0.179	0	1
Y85-88*SAT	Interaction term between Y85-88 and SAT - defined under student variables	36.001	194.366	0	1359.5
Y85-88*HS	Interaction term between Y85-88 and High School - defined under student variables	2.713	14.857	0	99.434
	Instructor Variables				
Instructor	Instructor name used as a level, 1,377 instructors				
Gender	Gender of the instructor, male = 1 and female = 0	0.780	0.414	0	1
Prof	Equals 1 if the position at the time of instruction was professor, 0 otherwise (dropped to avoid perfect collinearity)	0.414	0.493	0	1
Assoc prof	Equals 1 if the position at the time of instruction was associate professor, 0 otherwise	0.212	0.408	0	1
Assist prof	Equals 1 if the position at the time of instruction was assistant professor, 0 otherwise	0.134	0.341	0	1
Lec Grad	Equals 1 if the position at the time of instruction was graduate student, 0 otherwise	0.101	0.301	0	1

Other	Equals 1 if the position at the time of instruction was other lecturer, 0 otherwise (includes visiting faculty, lecturers, non-graduate instructors)	0.140	0.347	0	1
AAU	Equals 1 if the university was AAU member at the time of the instructor's graduation (includes Canadian universities), 0 otherwise (dropped to avoid collinearity)	0.365	0.481	0	1
Non-AAU	Equals 1 if the university was not AAU member at the time of the instructor's graduation (includes Canadian universities), 0 otherwise	0.527	0.499	0	1
Foreign	Equals 1 if the instructor's terminal degree was from a non-US or non-Canadian university, 0 otherwise	0.019	0.137	0	1
	Student Variables				
Percent Male	Percentage of male students in the class	0.497	0.208	0	1
SAT	Class average of students' combined SAT scores	1101.751	71.280	721.667	1472.500
Load	Average number of credits students in the class are enrolled	13.959	0.762	2.828	20.667
High School rank	The average high school rank of students in the class, calculated as the percentile of students in the school that rank below the given student	79.424	8.375	12.903	99.857
Share	Share of students who enrolled in the class but did not receive an A – F grade for the class. Includes students who dropped beyond the initial drop date, received an incomplete grade, took the class pass / fail, or was dropped from the class by the dean's office divided by total students	0.032	0.050	0	0.982

Table A2. Estimated Coefficients of the GPA and StDev Models

	GPA			StDev		
Variable Name	Coefficient	Std. Error	P-value	Coefficient	Std. Error	P-value
Constant	1.552	0.093	0.000	1.862	0.066	0.000
Grade Inflation Variables						
Ln trend	0.026	0.007	0.000	0.019	0.005	0.000
Institutional Variables						
Afternoon	0.051	0.006	0.000	-0.045	0.005	0.000
Evening	0.047	0.018	0.010	-0.077	0.014	0.000
Meet 1	0.091	0.011	0.000	-0.035	0.009	0.000
Meet 2	0.003	0.008	0.653	0.003	0.006	0.654
Upper division	0.004	0.009	0.683	-0.055	0.007	0.000
Total Students	-0.002	0.0001	0.000	0.001	0.0001	0.000
High credit	-0.261	0.010	0.000	0.116	0.008	0.000
Semester	0.008	0.005	0.105	-0.006	0.004	0.129
Y85-88	0.727	0.232	0.002	-0.745	0.178	0.000
Y85-88*SAT	-0.0004	0.0002	0.065	0.001	0.0002	0.001
Y85-88*HS	-0.004	0.001	0.000	0.002	0.001	0.004
Instructor Variables						
Gender	-0.013	0.019	0.500	-0.016	0.011	0.167
Assoc prof	0.021	0.010	0.028	0.009	0.007	0.201
Assist prof	0.015	0.012	0.227	0.024	0.009	0.006
Lec Grad	0.038	0.019	0.041	0.016	0.012	0.205
Other	0.100	0.022	0.000	-0.032	0.014	0.026
Non-AAU	-0.022	0.011	0.044	0.009	0.008	0.242
Foreign	-0.092	0.037	0.013	0.044	0.025	0.081
Student Variables						
Percent Male	-0.181	0.019	0.000	0.137	0.015	0.000
SAT	0.001	0.0001	0.000	-0.001	0.0001	0.000
Load	0.012	0.003	0.000	-0.016	0.003	0.000
High School rank	0.004	0.0004	0.000	-0.003	0.0003	0.000
Share	-1.046	0.049	0.000	0.638	0.038	0.000

		Random-Effects Parameters		
Department	0.026	0.010	0.007	0.003
Instructor	0.074	0.004	0.020	0.001
Residual	0.082	0.001	0.051	0.001
		Overall Model Fit		
Wald Test	$\chi^2 = 7577.47$	P-value 0.000	$\chi^2 = 3957.16$	P-value = 0.000
Likelihood Ratio vs. Linear Model	$\chi^2 = 9046.42$	P-value 0.000	$\chi^2 = 4272.74$	P-value = 0.000

Table A3. Variable Description for Departments' Model

Variable Name	Description
GPA	Unweighted class mean GPA
Ln trend	Natural logarithm of trend as given by semester
Morning	Equals 1 if class starts before 12:01, 0 otherwise
Afternoon	Equals 1 if class starts between 12:01 to 15:59, 0 otherwise (dropped to avoid perfect collinearity)
Meet 1	Equals 1 if the class meets once per week – usually class duration is 2.5 hours for a three-credit class, 0 otherwise (dropped to avoid perfect collinearity)
Meet 2	Equals 1 if the class meets twice per week – usually class duration is 75 minutes for a three-credit class, 0 otherwise
Meet 3	Equals 1 if the class meets three times per week – usually class duration is 50 minutes for a three-credit class, 0 otherwise
Lower division	Equals 1 if the class is listed as a 100 or 200 level class, 0 otherwise (dropped to avoid perfect collinearity)
Upper division	Equals 1 if the class is listed as a 300 or 400 level class, 0 otherwise
Total students	Number of students receiving a grade A – F and no grades (see share below) in the class
Low credit	Equals 1 if the class is 1 or 2 credit hours, 0 otherwise – very few classes are 2 credits (dropped to avoid perfect collinearity)
High credit	Equals 1 if the class is 3 credit hours or more, 0 otherwise very few classes have more than3 credits
Instructor	Instructor name used as a level, 1,377 instructors

Instructor gender	Gender of the instructor, male = 1 and female = 0
Professor	Equals 1 if the position at the time of instruction was professor, 0 otherwise (dropped to avoid perfect collinearity)
Associate prof	Equals 1 if the position at the time of instruction was associate professor, 0 otherwise
Assistant prof	Equals 1 if the position at the time of instruction was assistant professor, 0 otherwise
Lecturer graduate	Equals 1 if the position at the time of instruction was graduate student, 0 otherwise
Other lecture	Equals 1 if the position at the time of instruction was other lecturer, 0 otherwise (includes visiting faculty, lecturers, non-graduate instructors)
AAU	Equals 1 if the university was AAU member at the time of the instructor's graduation (includes Canadian universities), 0 otherwise
Student gender	Percentage of male students in the class
SAT	Class average of students' SAT math scores
Student load	Average number of credits students in the class are enrolled
HS percentile	The average high school rank of students in the class, calculated as the percentile of students in the school that rank below the given student
Share/no grade	Share of students who enrolled in the class but did not receive an A – F grade for the class. Includes students who dropped beyond the initial drop date, received an incomplete grade, took the class pass / fail, or was dropped from the class by the dean's office divided by total students

Table A4. Variable Mean Values in 1989-2003 and 2004-2019 and t-tests for Differences in Mean Values.

		Social Sciences			Plant Oriented			Animal Oriented			Other		
Variable	Period	AGEC	ALEC	RPTS	HORT	PLPA	SCSC	ANSC	POSC	BAEN	BICH	ENTO	WFSC
GPA	1	2.987	3.609	2.911	3.414	3.009	3.043	3.307	3.213	3.262	3.015	3.150	3.095
	2	3.098	3.474	3.073	3.435	3.250	3.100	3.482	3.280	3.179	3.085	3.261	2.998
	Diff.	0.111*	-0.135*	0.162*	0.021	0.241*	0.057	0.175*	0.067	-0.083*	0.070*	0.111*	-0.097*
Institutional Variables													
Morning	1	0.657	0.726	0.691	0.747	0.480	0.702	0.815	0.621	0.805	0.622	0.819	0.685
	2	0.556	0.548	0.518	0.564	0.384	0.602	0.714	0.647	0.668	0.560	0.517	0.591
	Diff.	-0.101*	-0.178*	-0.173*	-0.183*	-0.096	-0.100*	-0.101*	0.026	-0.137*	-0.062*	-0.302*	-0.094*
After-noon	1	0.343	0.272	0.309	0.253	0.520	0.298	0.185	0.379	0.195	0.378	0.181	0.315
	2	0.444	0.452	0.482	0.436	0.616	0.398	0.286	0.353	0.332	0.440	0.483	0.409
	Diff.	0.101*	0.178*	0.173*	0.183*	0.096	0.100*	0.101*	-0.026	0.137*	0.062*	0.302*	0.094*
Meet 1	1	0.000	0.444	0.170	0.386	0.440	0.284	0.448	0.316	0.276	0.195	0.046	0.172
	2	0.068	0.349	0.063	0.472	0.530	0.232	0.502	0.226	0.103	0.229	0.295	0.287
	Diff.	0.068*	-0.095*	-0.107*	0.086*	0.090	-0.052	0.054*	-0.090*	-0.173*	0.034	0.249*	0.115*
Meet 2	1	0.579	0.534	0.439	0.593	0.353	0.369	0.362	0.353	0.514	0.419	0.858	0.737
	2	0.634	0.530	0.581	0.511	0.315	0.437	0.300	0.380	0.620	0.460	0.513	0.665
	Diff.	0.055*	-0.004	0.142*	-0.082*	-0.038	0.068*	-0.062*	0.027	0.106*	0.041	-0.345*	-0.072*
Meet 3	1	0.421	0.022	0.390	0.021	0.207	0.348	0.190	0.331	0.210	0.386	0.096	0.090
	2	0.298	0.122	0.355	0.017	0.156	0.331	0.198	0.395	0.277	0.311	0.192	0.048
	Diff.	-0.123*	0.100*	-0.035	-0.004	-0.051	-0.017	0.008	0.064	0.067*	-0.075*	0.096*	-0.042*
Upper division	1	0.891	0.858	0.756	0.528	0.893	0.870	0.660	0.599	0.707	0.963	0.552	0.856
	2	0.866	0.876	0.847	0.653	0.877	0.842	0.662	0.786	0.880	0.958	0.685	0.922
	Diff.	-0.025	0.018	0.091*	0.125*	-0.016	-0.028	0.002	0.187*	0.173*	-0.005	0.133*	0.066*
Total students	1	62.195	55.688	43.32	45.514	36.507	45.576	63.775	39.342	31.185	62.711	41.915	41.846
	2	69.773	40.612	46.967	48.201	41.414	47.063	62.718	40.282	35.971	48.602	53.942	34.300
	Diff.	7.5783*	-15.080*	3.647*	2.687	4.907	1.488	-1.057	0.940	4.786*	-14.110*	12.027*	-7.546*
	1	1.000	0.490	0.871	0.622	0.673	0.7143	0.519	0.695	0.726	0.734	0.957	0.836

High credit	2	0.983	0.808	0.957	0.650	0.526	0.801	0.502	0.724	0.902	0.720	0.765	0.853
	Diff.	-0.017*	0.318*	0.086*	0.028	-0.147*	0.087*	-0.017	0.029	0.176*	-0.014	-0.192*	0.017
						Instructor Variables							
Instruct. gender	1	0.951	0.611	0.782	0.764	0.627	0.967	0.859	0.960	0.890	0.779	0.993	0.912
	2	0.914	0.461	0.694	0.846	0.798	0.887	0.786	0.905	0.800	0.695	0.798	0.860
	Diff.	-0.037*	-0.150*	-0.088*	0.082*	0.171*	-0.080*	-0.073*	-0.055*	-0.090*	-0.083*	-0.195*	-0.052*
Professor	1	0.546	0.195	0.294	0.376	0.340	0.484	0.446	0.386	0.499	0.478	0.918	0.539
	2	0.654	0.148	0.214	0.506	0.629	0.489	0.437	0.309	0.478	0.378	0.495	0.585
	Diff.	0.108*	-0.047*	-0.079*	0.131*	0.289*	0.004	-0.009	-0.077*	-0.020	-0.100*	-0.423*	0.046
Assistant prof	1	0.138	0.230	0.235	0.093	0.093	0.029	0.129	0.081	0.123	0.104	0.050	0.132
	2	0.109	0.294	0.130	0.034	0.126	0.130	0.112	0.315	0.122	0.116	0.165	0.090
	Diff.	-0.029	0.064*	-0.105	-0.058*	0.032	0.101*	-0.016	0.234*	-0.001	0.012	0.115*	-0.042*
Assoc. prof	1	0.194	0.227	0.205	0.117	0.240	0.327	0.167	0.346	0.257	0.197	0.028	0.178
	2	0.091	0.249	0.243	0.105	0.215	0.294	0.260	0.252	0.241	0.166	0.194	0.224
	Diff.	-0.102*	0.021	0.038	-0.012	-0.025	-0.033	0.093*	-0.093*	-0.016	-0.031	0.166*	0.046
Lecture graduate	1	0.068	0.285	0.138	0.184	0.313	0.077	0.185	0.074	0.019	0.040	0.004	0.072
	2	0.054	0.165	0.206	0.057	0.000	0.003	0.140	0.077	0.019	0.021	0.049	0.057
	Diff.	-0.014	-0.119*	0.068*	-0.126*	-0.313*	-0.074*	-0.044*	0.004	0.000	-0.019*	0.045*	-0.016
Other lecturer	1	0.055	0.063	0.129	0.231	0.012	0.083	0.074	0.114	0.102	0.181	0.000	0.078
	2	0.091	0.144	0.208	0.297	0.030	0.084	0.051	0.047	0.139	0.319	0.097	0.044
	Diff.	0.037*	0.081*	0.079*	0.066*	0.028	0.001	-0.023*	-0.066*	0.038	0.138*	0.097*	-0.034*
Non-AAU	1	0.595	0.526	0.566	0.512	0.827	0.542	0.753	0.691	0.843	0.513	0.566	0.816
	2	0.385	0.561	0.586	0.554	0.659	0.574	0.549	0.472	0.711	0.476	0.348	0.711
	Diff.	-0.210*	0.035	0.020	0.042	-0.168*	0.032	-0.204*	-0.219*	-0.132*	-0.037	-0.218*	-0.105*
						Student Variables							
Student gender	1	0.650	0.516	0.502	0.398	0.555	0.739	0.493	0.656	0.802	0.445	0.591	0.561
	2	0.648	0.424	0.418	0.358	0.461	0.706	0.320	0.557	0.781	0.400	0.411	0.508
	Diff.	-0.003	-0.092*	-0.084*	-0.040*	-0.093*	-0.033*	-0.173*	-0.099*	-0.022*	-0.045*	-0.180*	-0.054*
SAT	1	536.455	528.940	528.022	541.225	549.957	547.436	546.728	525.742	563.844	603.335	551.364	554.240
	2	550.955	527.968	543.309	558.052	569.051	549.944	560.975	532.083	592.254	623.538	576.754	568.296

	Diff.	14.500*	-0.973	15.287*	16.827*	19.094*	2.508	14.247*	6.341*	28.410*	20.204*	25.391*	14.056*
Student load	1	14.147	14.223	13.762	13.974	13.918	14.016	14.063	14.323	14.227	14.127	13.974	13.934
	2	13.678	14.114	14.129	14.019	13.955	13.924	13.854	14.286	14.031	13.853	14.019	14.084
	Diff.	-0.468*	-0.109*	0.367*	0.045*	0.038	-0.092*	-0.209*	-0.037	-0.196*	-0.274*	0.045	0.150*
HS percent.	1	76.302	76.146	73.970	78.296	79.295	78.355	81.741	73.857	81.949	88.209	80.167	79.897
	2	73.042	73.031	73.324	78.867	80.296	74.064	83.701	74.391	81.056	87.655	83.949	80.583
	Diff.	-3.260*	-3.115*	-0.646	0.571	1.001	-4.291*	1.960*	0.534	-0.893	-0.554*	3.782*	0.686
Share/no grade	1	0.039	0.019	0.043	0.032	0.040	0.041	0.026	0.027	0.021	0.063	0.030	0.042
	2	0.026	0.019	0.042	0.027	0.019	0.031	0.022	0.027	0.017	0.051	0.028	0.037
	Diff.	-0.014*	0.000	-0.001	-0.005*	-0.021	-0.010*	-0.004*	0.000	-0.004*	-0.012	-0.003	-0.005

* significant at 0.05 and smaller (p-value < 0.05).

Table A5. Mixed Effect Model Estimated Coefficients for Period 1 (Years 1989 - 2003)

	Social Sciences			Plant Oriented			Animal Oriented			Other		
Variable	AGEC	ALEC	RPTS	HORT	PLPA	SCSC	ANSC	POSC	BAEN	BICH	ENTO	WFSC
Trend	0.006*	-0.004	-0.002	0.002	-0.0003	0.005*	0.005*	-0.012*	0.008*	0.001	-0.001	-0.002
	(0.001)	(0.002)	(0.003)	(0.002)	(0.006)	(0.002)	(0.001)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Institution Variables												
Morning	-0.060*	-0.001	-0.097*	0.073*	-0.030	-0.100*	-0.074*	-0.208*	-0.037	0.013	-0.191*	-0.081
	(0.020)	(0.030)	(0.037)	(0.031)	(0.110)	(0.050)	(0.025)	(0.057)	(0.038)	(0.022)	(0.063)	(0.045)
Meet 2	0.051*	-0.142*	-0.111	-0.177	-0.061	-0.471	-0.277*	-0.405*	-0.158*	-0.187*	-0.347	0.018
	(0.024)	(0.034)	(0.084)	(0.140)	(0.118)	(0.286)	(0.033)	(0.199)	(0.048)	(0.058)	(0.251)	(0.154)
Meet 3	n/a	-0.391*	-0.087	-0.119	-0.627*	-0.789*	-0.254*	-0.205	-0.056	-0.206*	-0.146	0.101
		(0.086)	(0.088)	(0.163)	(0.150)	(0.292)	(0.042)	(0.201)	(0.062)	(0.064)	(0.263)	(0.163)
Upper division	-0.005	-0.017	0.111	0.018	-0.205	-0.223*	0.146*	-0.102	0.134*	0.108	-0.131	-0.054
	(0.058)	(0.043)	(0.047)	(0.035)	(0.206)	(0.093)	(0.035)	(0.058)	(0.047)	(0.062)	(0.090)	(0.061)
Total students	-0.001*	-0.002*	-0.004*	-0.002*	-0.008*	-0.002*	-0.002*	-0.001*	-0.003*	-0.003*	-0.003*	-0.002*
	(0.0003)	(0.0003)	(0.001)	(0.0003)	(0.002)	(0.001)	(0.0002)	(0.001)	(0.001)	(0.0003)	(0.001)	(0.001)
High credit	n/a	-0.152*	-0.069	-0.165	-0.139	0.171	-0.033	-0.018	-0.231*	-0.493*	-0.098	-0.700*
		(0.035)	(0.094)	(0.142)	(0.128)	(0.293)	(0.033)	(0.203)	(0.052)	(0.056)	(0.297)	(0.156)
Instructor Variables												
Instructor gender	-0.0002	-0.014	-0.022	-0.050	-0.101	-0.004	-0.036	-0.147	0.166	0.074	☹	-0.042
	(0.134)	(0.055)	(0.092)	(0.066)	(0.090)	(0.163)	(0.067)	(0.140)	(0.157)	(0.072)		(0.116)
Assistant prof	0.050	0.060	0.081	0.006	0.036	-0.263*	0.069	-0.099	0.001	-0.008	-0.287*	-0.046
	(0.050)	(0.063)	(0.091)	(0.068)	(0.141)	(0.107)	(0.047)	(0.100)	(0.097)	(0.056)	(0.133)	(0.081)
Associate prof	0.050	-0.051	0.064	-0.036	0.026	0.078	0.058	0.056	-0.157	-0.005	-0.086	0.042
	(0.037)	(0.049)	(0.078)	(0.050)	(0.115)	(0.059)	(0.033)	(0.074)	(0.086)	(0.039)	(0.102)	(0.056)
Lecturer graduate	-0.103	0.141*	0.201	0.070	-0.124	-0.004	0.091	0.188	-0.071	-0.139	☹	0.006
	(0.778)	(0.068)	(0.114)	(0.076)	(0.146)	(0.150)	(0.059)	(0.117)	(0.192)	(0.097)		(0.096)

Other lecturer	-0.065	-0.006	0.204	0.180	☹	0.492*	0.153	0.310	-0.140	-0.049	n/a	0.224
	(0.136)	(0.090)	(0.130)	(0.114)		(0.193)	(0.083)	(0.171)	(0.162)	(0.100)		(0.149)
Non - AAU	0.129*	-0.123*	0.024	0.013	0.078	-0.073	-0.089	0.013	0.098	-0.049	-0.017	-0.044
	(0.064)	(0.054)	(0.098)	(0.058)	(0.135)	(0.121)	(0.064)	(0.094)	(0.112)	(0.062)	(0.105)	(0.088)
Student Variables												
Student gender	-0.575*	-0.299*	-0.130	-0.341*	-0.014	0.069	-0.136*	0.335*	-0.421*	-0.132	-0.110	-0.236*
	(0.100)	(0.112)	(0.147)	(0.087)	(0.236)	(0.114)	(0.067)	(0.172)	(0.112)	(0.080)	(0.114)	(0.107)
SAT	0.003*	0.003*	-0.0003	0.002*	0.0005	0.001	0.002*	0.001	0.001	0.005*	0.003*	-0.0001
	(0.001)	(0.001)	(0.001)	(0.0004)	(0.001)	(0.001)	(0.0004)	(0.001)	(0.001)	(0.0004)	(0.001)	(0.001)
Student load	0.037*	0.009	-0.022	0.001	-0.090	0.040	0.035*	-0.062*	0.007	0.039*	0.039*	0.026
	(0.018)	(0.016)	(0.019)	(0.013)	(0.050)	(0.024)	(0.112)	(0.020)	(0.018)	(0.017)	(0.016)	(0.024)
HS percentile	0.005*	0.003	-0.001	0.002	0.008*	0.001	0.008*	0.007*	0.006*	0.006*	0.006*	0.003
	(0.001)	(0.002)	(0.002)	(0.002)	(0.004)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
Share/no grade	-1.317*	-1.314*	-2.526*	-0.558*	-0.411	-1.031*	-0.808*	-1.963*	-1.602*	-0.695*	-0.104	-0.884*
	(0.212)	(0.354)	(0.358)	(0.261)	(0.613)	(0.270)	(0.239)	(0.546)	(0.383)	(0.174)	(0.405)	(0.296)
Random-Effect Parameters												
Instructor	0.569*	0.016*	0.073*	0.029*	0.041*	0.089*	0.045*	0.020*	0.066*	0.071*	0.050*	0.049*
Residual	0.048*	0.043*	0.107*	0.056*	0.075*	0.076*	0.055*	0.082*	0.072*	0.053*	0.054*	0.076*

* Significant at 0.05 or smaller (p-value < 0.05). Standard errors in parenthesis below the estimated coefficients.

When the number of observations is fewer than five observations (marked as ☹) the variable is removed for confidentiality reasons and because conclusions drawn would be suspect.

Table A6. Mixed Effect Model Estimated Coefficients for Period 2 (Years 2004 - 2019)

	Social Sciences			Plant Oriented			Animal Oriented		Other			
Variable	AGEC	ALEC	RPTS	HORT	PLPA	SCSC	ANSC	POSC	BAEN	BICH	ENTO	WFSC
Trend	0.006*	0.005*	0.004*	0.006*	0.008*	0.006*	0.001	0.008*	0.005*	-0.004*	0.005	-0.001
	(0.001)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.001)	(0.003)	(0.002)	(0.001)	(0.003)	(0.002)
Institution Variables												
Morning	-0.010	-0.054*	-0.034	-0.001	-0.025	-0.167*	-0.011	-0.200*	0.024	0.065*	-0.002	0.032
	(0.020)	(0.019)	(0.026)	(0.026)	(0.055)	(0.046)	(0.020)	(0.041)	(0.037)	(0.024)	(0.037)	(0.038)
Meet 2	0.113*	0.073*	-0.051	0.152*	-0.142	0.215*	0.016	-0.369*	-0.103	-0.273*	0.027	-0.142*
	(0.058)	(0.024)	(0.072)	(0.047)	(0.091)	(0.073)	(0.026)	(0.069)	(0.059)	(0.033)	(0.071)	(0.055)
Meet 3	0.109	0.075*	-0.031	0.154	-0.077	0.132	-0.029	-0.202*	-0.114	-0.295*	0.103	0.045
	(0.060)	(0.031)	(0.075)	(0.088)	(0.122)	(0.077)	(0.033)	(0.082)	(0.067)	(0.042)	(0.081)	(0.094)
Upper division	0.076	0.044	-0.030	-0.019	-0.123	-0.172*	0.175*	-0.068	0.005	-0.091	0.187*	-0.264*
	(0.044)	(0.033)	(0.041)	(0.040)	(0.076)	(0.071)	(0.027)	(0.058)	(0.055)	(0.057)	(0.054)	(0.071)
Total students	-0.001*	-0.002*	-0.001	-0.002*	-0.003*	-0.001*	-0.001*	-0.002*	-0.006*	-0.001*	-0.001*	-0.004*
	(0.0002)	(0.0004)	(0.001)	(0.0002)	(0.001)	(0.0005)	(0.0001)	(0.001)	(0.001)	(0.0003)	(0.0004)	(0.001)
High credit	-0.478*	-0.251*	-0.187*	-0.382*	-0.399*	-0.352*	-0.228*	0.003	-0.406*	-0.286*	-0.219*	-0.352*
	(0.084)	(0.025)	(0.088)	(0.051)	(0.080)	(0.071)	(0.025)	(0.071)	(0.062)	(0.036)	(0.076)	(0.062)
Instructor Variables												
Instructor gender	-0.096	0.126*	-0.190*	0.009	-0.079	0.187	-0.094	-0.096	-0.061	-0.081	-0.083	-0.171
	(0.096)	(0.058)	(0.076)	(0.112)	(0.079)	(0.144)	(0.060)	(0.165)	(0.095)	(0.082)	(0.150)	(0.133)
Assistant prof	0.106	0.071	0.034	-0.017	0.247*	0.166	-0.021	0.063	0.037	-0.089	-0.122	-0.068
	(0.064)	(0.059)	(0.073)	(0.095)	(0.085)	(0.096)	(0.049)	(0.078)	(0.083)	(0.056)	(0.086)	(0.085)
Associate prof	0.041	0.006	-0.035	-0.001	0.063	0.142	0.002	-0.042	0.024	-0.043	-0.167*	-0.106
	(0.047)	(0.052)	(0.055)	(0.060)	(0.066)	(0.091)	(0.033)	(0.064)	(0.066)	(0.048)	(0.071)	(0.068)
Lecturer graduate	0.065	0.120	0.073	0.223	n/a	☹	0.137*	- 0.063	-0.198	-0.150	-0.036	0.050
	(0.074)	(0.073)	(0.088)	(0.139)			(0.065)	(0.099)	(0.153)	(0.126)	(0.185)	(0.128)

Other lecturer	0.107	0.121	0.247*	0.080	0.132	0.481*	0.167	0.132	-0.123	0.005	0.098	-0.165
	(0.072)	(0.080)	(0.102)	(0.129)	(0.220)	(0.213)	(0.091)	(0.222)	(0.116)	(0.108)	(0.176)	(0.157)
Non - AAU	-0.023	-0.058	0.057	0.139	-0.033	-0.136	-0.069	-0.085	0.087	-0.037	0.078	-0.122
	(0.047)	(0.037)	(0.064)	(0.091)	(0.094)	(0.087)	(0.042)	(0.111)	(0.069)	(0.046)	(0.113)	(0.082)
Student Variables												
Student gender	-0.296*	-0.434	-0.265*	-0.053	-0.364*	-0.438*	-0.245*	0.050	-0.329*	-0.148*	-0.043	-0.493*
	(0.095)	(0.070)	(0.095)	(0.092)	(0.154)	(0.116)	(0.067)	(0.128)	(0.109)	(0.072)	(0.132)	(0.101)
SAT	0.003*	0.001*	-0.0005	0.001*	0.001	0.001	0.002*	-0.0003	0.001*	0.004*	0.004*	-0.0004
	(0.001)	(0.0004)	(0.001)	(0.001)	(0.001)	(0.001)	(0.0003)	(0.001)	(0.001)	(0.0004)	(0.001)	(0.001)
Student load	-0.018	0.023	0.028	0.004	-0.043	0.015	0.005	0.010	-0.090*	-0.015	0.051*	0.005
	(0.016)	(0.013)	(0.017)	(0.017)	(0.027)	(0.020)	(0.012)	(0.026)	(0.020)	(0.013)	(0.022)	(0.018)
HS percentile	0.005*	0.005*	0.001	0.008*	0.008*	0.001	0.007*	0.007*	0.009*	0.015*	0.007*	0.004
	(0.002)	(0.001)	(0.017)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.018)
Share/no grade	-1.986*	-2.081*	-0.054	-0.677*	-1.825*	-0.317	-1.106*	-0.279*	-1.110*	-0.842*	-0.977*	-0.891*
	(0.271)	(0.248)	(0.108)	(0.192)	(0.564)	(0.299)	(0.192)	(0.331)	(0.389)	(0.148)	(0.359)	(0.291)
Random-Effect Parameters												
Instructor	0.060*	0.058*	0.095*	0.068*	0.032*	0.125*	0.041*	0.087*	0.085*	0.085*	0.086*	0.097*
Residual	0.049*	0.067*	0.061*	0.064*	0.068*	0.071*	0.064*	0.058*	0.087*	0.069*	0.093*	0.072*

* Significant at 0.05 or smaller (p-value < 0.05). Standard errors in parenthesis below the estimated coefficients.

When the number of observations is fewer than five observations (marked as ☹) the variable is removed for confidentiality reasons and because conclusions drawn would be suspect.

Table A7. Outcome Categories

Categories of Outcomes	Outcomes
Production outcomes	- Yield - Technology adoption - Area under cultivation - Technical efficiency - Crop diversification - Product quality
Market reforms	- Market access - Agricultural profits - Awareness / knowledge on market demand and prices - Percent of crop sold - Cross-border trade - Inputs availability
Product movement in space and time	- Produce procurement services - Infrastructure (roads and facilities) - Food waste management - Food safety management
Risk management	- Yield risk - Price or market risk - Contract risk - Climate change risk
Welfare	- Family income - Food security and food access - Nutritional security - Poverty reduction - Education and healthcare - Employment conditions and opportunities - Environmental conservation or pollution reduction - Savings and consumption

Table A8. Categories of Intervention for Each Stage of the AVC.

Stages of the Value Chain	Categories of Interventions
Input supply stage	- Input distribution network and management - Information sharing and knowledge management - Institutional credit and subsidies - Land tenure and land improvement - Water management
Production, harvest, post-harvest stage	- Crop management services - Dairy, meat, and livestock related services
Transport and storage stage	- Transport capacity (roads, refrigerated vehicles) and modernized storage
Processing stage	- Value addition services - Product traceability and food safety services
Market access	- Contract farming and industry tie-up - Certification along value chain - Commodity markets and trading platforms - Market information networks and channels
Governance and inclusion	- Interventions around gender, poverty, and social issues - Value chain coordination - Trade policies

APPENDIX B – SUPPLEMENTARY MATERIALS

Appendix B1. Eligibility or Inclusion/Exclusion Criteria for Evidence Gap Map Studies.

	Inclusion	Exclusion
P: Population	- Agriculture Value chains whose production component is done in developing countries (low, middle, and upper-middle income countries according to World Bank classification 2018). - Value chain stakeholders (farmers/smallholder farmers/peasants/farmer producer groups, Farmer's cooperatives traders, processors/processing companies, retailers/wholesalers, input suppliers, end users). - External stakeholders (governments, multilateral, international and national NGOs, and UN agencies).	- Non-agricultural production (other than cultivating the soil, growing crops, raising livestock, and fish-farming in ponds) which includes wildlife use and extraction, forest products extraction, wild fisheries, commercial fisheries. - Urban agriculture.
I: Intervention	- Policies adopted by governments, multilateral, international and national NGOs, and UN agencies to enhance agricultural value chains. For example: trade liberalization and agreement policy, pricing policy, agricultural subsidies, market access policy, credit policy, FDI Investment policy/donor policies, conservation set-aside policy, agriculture extension and technology adoption policy, partner engagement policy, road and transport policy, land tenure policy, water policy/irrigation policy, forest/water resource policy. *Note*: It includes governmental policies targeted to one stakeholder group in the value chain. - Policies within the value chain adopted by the stakeholders with or without the direct participation/intervention of governments. For example: contract farming, certifications, Agriculture	- Evaluation of the normal course of action of an organization. Example: impact of credit access from regular/formal banking services, cooperative membership, unconditional cash transfers. - Policies on natural resources allocation or related subjects that had an unintended effect on value chains. - Since policies are usually long-term courses of action, impact evaluations of projects within programs or isolated projects will not be considered.

	information services, supermarket place and infrastructure policy, other market mechanisms of "self" adjustment.	
C: Comparison	- Policy/services as usual - Policy/program vs another Policy/program - Policy/program vs doing nothing or wait-list control - Early-vs-late comparison in the implementation of a policy/program	- Studies presenting only a diagnosis of a Value Chain. - Historical analysis without comparison. - Studies analyzing single pre- and pos t-test data without comparison.
O: Outcomes	- A preliminary list of possible groups of outcomes includes (but not limited to): - Input flow: access for small-scale farmers, small-scale farmers labor decisions - Product flow (involves changes in supply and demand, elasticity of supply, elasticity of demand): cold chain management, demand management, food processing, food safety and quality, harvesting, logistics, marketing channel, packing system, post-harvest loss, traceability. - Financial flow - Market prices: price transmission, pricing - Performance: risk management, information flow, performance measurements, innovations, perishability, procurement model, waste management. Enablers/barriers: ICT, competitiveness, knowledge management, among others.	
S: Study design	- Effectiveness studies (primary concerned with the impact). - Implementation studies (focused on how and why policies have a particular impact, including failure in achieving the impact). - - For studies providing quantitative evidence: evaluations* that use randomized designs, quasi-RCT, natural experiments, or methods to identify	- Simulation studies, willingness to pay, and hypothetical experiment studies. - Summaries, overviews, or reports of the projects supported by governments or NGO with no detailed reports in the impact evaluation. - Qualitative studies using

	causation among self-elected groups (pre-and-post test data with comparison, multiple pre- and post- test data without comparison, cross-sectional with comparison, post-test studies using instrumental variables). - For studies providing qualitative evidence: studies using rigorously qualitative research methods such as discourse analysis techniques, thematic analysis techniques, grounded theory, phenomenological studies and ethnographic methods, or any combination of them (triangulation approaches). Computing correlation and statistical tests or considering alternative/contradictory evidence. - Mixed-methods (using any combination of the quantitative and qualitative evidence listed above) will be included. - Systematic reviews - Systematic review protocols	content analysis or narrative analysis. - Policy announcements, descriptions, and articles.
Additional criteria	- Written in English. - Studies published since 2000	- Documents with no full-text reporting in English. -

What models of collaboration among local and international actors, including donors, private sector partners, academic institutions, and NGOs, are effective in supporting policy change?

Focus on	- Governance (type) - Specific role of stakeholders in supporting functions and mechanisms (dynamics): AVC collaboration, Challenges in AVC, Co-ordination in AVC, Globalization of AVC, Linkages between drivers of AVC, Material and information flow, Performance measurements, post-harvest innovation, Role of Government (regulatory, support), sourcing strategies, structure of AVC, sustainability of AVC.

Appendix B2. List of Selected Studies.

Abate, G., T. Bernard, M. Regassa, and B. Minten. 2021. Improving Coffee Productivity in Ethiopia: The Impact of a Coffee Tree Rejuvenation Training Program on Stumping. *International Food Policy Research Institute (IFPRI)*. Discussion Paper 02023. https://www.ifpri.org/publication/improving-coffee-productivity-ethiopia-impact-coffee-tree-rejuvenation-training-program. (Accessed July 15, 2022).

Abdoulaye, T., P. Amaza, A. Olanrewaju, and J. Ellis-Jones. 2012. Promoting the Adoption of Innovations Through Participatory Approaches: Example from Northern Nigeria. *Tropicultura* 30(3): 155-160.

Abokyi, E., H. Folmer, and K. Asiedu. 2018. Public Buffer Stocks as Agricultural Output Price Stabilization Policy in Ghana. *Agriculture and Food Security 7*(1): 1-12.

Adékambi, S. 2015. The Roles of Exploration and Exploitation in the Export Market Integration of Beninese Producers at the Base of the Pyramid. Wageningen University and Research. https://edepot.wur.nl/333126. (Accessed July 16, 2022).

Agyemang, S., T. Ratinger, and S. Ahado. 2020. Has Microcredit Boosted Poultry Production in Ghana? *Agricultural Finance Review 80*(2): 135-152.

Ahmed, M., S. Sultana, S. Halim, and M. Islam. 2016. Marketing Systems of Freshwater Prawns in Three Coastal Districts of Bangladesh. *Aquaculture Economics & Management 20*(3): 272-282.

Andani, A., A. Moro, and G. Issahaku. 2020. Fertilizer Subsidy Policy and Smallholder Farmers Crop Productivity: The Case of Maize Production in North-Eastern Ghana. *Journal of Agricultural Extension and Rural Development 12*(2): 18-25.

Antwi, K. 2018. Essays in Rural Household Food Security and Nutrition in Ghana. Dissertation, Texas Tech University. https://ttu-ir.tdl.org/handle/2346/73830. (Accessed July 16, 2022).

Ao, G., Q. Liu, L. Qin, M. Chen, S. Liu, and W. Wu. 2021. Organization Model, Vertical Integration, and Farmers' Income Growth: Empirical Evidence from Large-Scale Farmers in Lin'an, China. *Plos One* 16(6): e0252482.

Arbeláez, M., S. Higuera, R. Steiner, and S. Zuluaga. 2018. Protecting "Sensitive" Agricultural Products in Colombia: A Political Economy Approach. *Coyuntura Economica: Investigaciao Economica y Social* XLVIII (1 y 2): 153-205.

Awe, F. 2020. Impact of Access to Credit on Farm Output Among Agroforestry Farmers in Oyo State, Nigeria. *Nigeria Agricultural Journal 51*(3): 91-99.

Aye, M. 2017. Rice contract Farming System and its Impact on Smallholder Livelihood in Myanmar. Margraf Verlag. ISBN: 9783823617402.

Balié, J., D. Del Prete, E. Magrini, P. Montalbano, and S. Nenci. 2019. Does Trade Policy Impact Food and Agriculture Global Value Chain Participation of Sub-Saharan African Countries? *American Journal of Agricultural Economics* 101(3): 773-789.

Belissa, T., R. Lensink, and A. Winkel. 2020. Effects of Index Insurance on Demand and Supply of Credit: Evidence from Ethiopia. *American Journal of Agricultural Economics 102*(5): 1511-1531.

Benjamin, E., O. Ola, H. Lang, and G. Buchenrieder. 2021. Public-Private Cooperation and Agricultural Development in Sub-Saharan Africa: A Review of Nigerian Growth Enhancement Scheme and E-Voucher Program. *Food Security 13*: 129-140.

Bernstein, J., N. Johnson, and A. Arslan. 2019. Meta-Evidence Review on the Impacts of Investments in Agricultural and Rural Development on Sustainable Development Goals 1 and 2. *IFAD Research Series 38*: 1-80.

Biggeri, M., F. Burchi, F. Ciani, and R. Herrmann. 2018. Linking Small-Scale Farmers to the Durum Wheat Value Chain in Ethiopia: Assessing the Effects on Production and Wellbeing. *Food Policy 79*: 77-91.

Bodnár, F. 2011. Improving Food Security: A Systematic Review of the Impact of Interventions in Agricultural Production, Value Chains, Market Regulation, and Land Security. The Hague (Netherlands) MFA. https://agris.fao.org/agris-search/search.do?recordID=XF2015044685. (Accessed July 19, 2022).

Bonan, J., H. Kazianga, and M. Mendola. 2020. Agricultural Transformation and Farmers' Expectations: Experimental Evidence from Uganda. Global Labor Organization (GLO) Discussion Paper, No. 456. https://www.econstor.eu/bitstream/10419/213041/1/GLO-DP-0456.pdf (Accessed May 26, 2023).

Booker, A., D. Johnston, and M. Heinrich. 2016. The Welfare Effects of Trade in Phytomedicines: A Multi-Disciplinary Analysis of Turmeric Production. *World Development 77*: 221-230.

Brabez, F. and M. Benmehaia. 2018. Vertical Relationships and Food Supply Chain Coordination: The Case of Processing Tomato Sector in Algeria. *New Medit: Mediterranean Journal of Economics, Agriculture, and Environment 17*(2): 3-14.

Brüntrup, M. 2011. The Comprehensive Africa Agriculture Development Programme (CAADP)–An Assessment of a Pan-African Attempt to Revitalize Agriculture. *Quarterly Journal of International Agriculture 50*(892-2016-65189): 79-106.

Cai, Y. and L. Yu. 2018. Rural Household Participation in and Satisfaction with Compensation Programs Targeting Farmland Preservation in China. *Journal of Cleaner Production 205*: 1148-1161.

Chandio, A., Y. Jiang, F. Wei, and X. Guangshun. 2018. Effects of Agricultural Credit on Wheat Productivity of Small Farms in Sindh, Pakistan: are Short-Term Loans Better? *Agricultural Finance Review* 78(5): 592-610.

Chen, X., G. Qiu, L. Wu, G. Xu, J. Wang, and W. Hu. 2017. Influential Impacts of Combined Government Policies for Safe Disposal of Dead Pigs on Farmer Behavior. *Environmental Science and Pollution Research* 24: 3997-4007.

Chibwana, C., M. Fisher, and G. Shively. 2012. Cropland Allocation Effects of Agricultural Input Subsidies in Malawi. *World Development* 40(1): 124-133.

Colen, L., M. Maertens, and J. Swinnen. 2012. Private Standards, Trade and Poverty: GlobalGAP and Horticultural Employment in Senegal. *The World Economy 35*(8): 1073-1088.

D'Hôtel, E. and P. Bosc. 2011. Neither State nor Market: The Influence of Farmers' Organizations on Agricultural Policies in Costa Rica. *Oxford Development Studies 39*(4): 469-485.

Dang, C. and V. Nguyen. 2019. Impacts of the Community Education on Farmers' Knowledge and Behavior in Pesticide Use Toward Pesticide Risk Reduction in Vietnam: a Case Study in Vegetable Production. *Journal of International Society for Southeast Asian Agricultural Sciences 25*(2): 155-165.

Dangour, A., S. Hawkesworth, B. Shankar, L. Watson, C. Srinivasan, E. Morgan, L. Haddad, and J. Waage. 2013. Can Nutrition be Promoted through Agriculture-Led Food Price Policies? A Systematic Review. *BMJ Open 3*(6): e002937.

Dannenberg, P. and G. Nduru. 2013. Practices in International Value Chains: the Case of the Kenyan Fruit and Vegetable Chain Beyond the Exclusion Debate. *Tijdschrift voor Economische en Sociale Geografie* 104(1): 41-56.

Delavallade, C. and S. Godlonton. 2023. Locking Crops to Unlock Investment: Experimental Evidence on Warrantage in Burkina Faso. *Journal of Development Economics* 160: 102959.

Díaz-Montenegro, J., E. Varela, and J. Gil. 2018. Livelihood Strategies of Cacao Producers in Ecuador: Effects of National Policies to Support Cacao Farmers and Specialty Cacao Landraces. *Journal of Rural Studies* 63: 141-156.

Dihel, N., A. Goswami, C. Hollweg, and A. Slany. 2018. How Does Participation in Value Chains Matter to African Farmers? *World Bank Policy Research Working Paper* (8506). https://openknowledge.worldbank.org/entities/publication/22de6df5-5b61-56da-a0e2-ce503fe8869d. (Accessed March 26, 2023).

Do, M. and S. Park. 2018. New Rural Development and Hierarchical Governance in Vietnam: Impacts of Government Support on Rural Households' Income Using a Hierarchical Linear Modelling. *Agris On-Line Papers in Economics & Informatics* 10(4): 3-15.

Donoso, G., O. Melo Contreras, and E. Negrete Allendes. 2010. Participation Determinants and Impact Assessment of the Institute of Agricultural Development (INDAP) Credit Program. *Ciencia e Investigación Agrarian* 37(2): 87-99.

Dougherty, J., R. Gallenstein, and K. Mishra. 2021. Impact of Index Insurance on Moral Hazard in the Agricultural Credit Market: Theory and Evidence from Ghana. *Journal of African Economies 30*(5): 418-446.

Ely, R., R. Parfitt, A. Carraro, and F. Ribeiro. 2019. Rural Credit and the Time Allocation of Agricultural Households: The Case of PRONAF in Brazil. *Review of Development Economics 23*(4): 1863-1890.

Esther, W. 2018. Innovation Development and Transfer by Agricultural Development Agencies: A Case Study of Cowpea IPM in Northern Ghana. *Agro-Science 17*(1): 7-19.

Fadeyi, O., B. Omojeso, and I. Ityokumbul. 2021. A Review of Microfinance Banks Effects on Smallholder Development in Nigeria. *African Journal of Agricultural Research 17*(9): 1249-1255.

Figueiredo, H., M. Meuwissen, and A. Oude Lansink. 2016. Evaluating Strategies for Honey Value Chains in Brazil Using a Value Chain Structure-Conduct-Performance (SCP) Framework. *International Food and Agribusiness Management Review* 19(3): 1-26.

Flifli, V., P. Okuneye, and D. Akerele. 2020. Lenders and Borrowers' Collaboration-Based Risk Mitigation Credit Market: Factors Influencing Access to Formal Credit in the Agricultural Sector in the Benin Republic. *Agricultural Finance Review 80*(2): 173-199.

Fuje, H. and H. Pullabhotla. 2020. Impact of Grain Trade Policies on Prices and Welfare. *Policy Research Working Paper Series 9436. The World Bank.* http://hdl.handle.net/10986/34631. (Accessed July 26, 2022).

Gaihre, S., J. Kyle, S. Semple, J. Smith, M. Subedi, and D. Marais. 2016. Type and Extent of Trans-Disciplinary Co-operation to Improve Food Security, Health and Household Environment in Low and Middle Income Countries: Systematic Review. *BMC Public Health 16*: 1-19.

Gao, Y., Y. Shu, H. Cao, S. Zhou, and S. Shi. 2021. Fiscal Policy Dilemma in Resolving Agricultural Risks: Evidence from China's Agricultural Insurance Subsidy Pilot. *International Journal of Environmental Research and Public Health 18*(14): 7577.

Garbero, A., D. Diatta, and M.Olapade. 2018. Impact Assessment-Agricultural Value Chains Support Project (PAFA): Senegal. *IFAD project report.* https://www.ifad.org/documents/38714170/41115672/SN_PAFA_IA+report.pdf/0deb15a4-9d3f-de7e-2b56-a2a753cf87dd?t=1557928407000. (Accessed May 26, 2023).

Garbero, A., P. Marion, and V. Brailovskaya. 2018. The Impact of the Adoption of CGIAR's Improved Varieties on Poverty and Welfare Outcomes: A Systematic Review. *IFAD Research Series* 33.

https://www.ifad.org/documents/38714170/40951886/Research+Series+33.pdf (Accessed May 26, 2023).

Getachew, M. and A. Engdawork. 2019. The Pursuit for Enhancing the Net Income and Sustainable Livelihood for Smallholder Farmers: The Case of Contract Farming in Oromia Regional State, Ethiopia. *Journal of Sustainable Development in Africa* 21(4): 146-177.

Ghins, L. and K. Pauw. 2018. The Impact of Markets and Policy on Incentives for Rice Production in Rwanda. *FAO Agricultural Development Economics Working Papers*, 03 Jul 2018, No.2018/02, DOI: https://doi.org/10.18356/4fbf2f35-en. (Accessed May 26, 2023).

González-Ramírez, M., V. Santoyo-Cortés, J. Arana-Coronado, and M. Muñoz-Rodríguez. 2020. The Insertion of Mexico into the Global Value Chain of Berries. *World Development Perspectives* 20: 1-11.

Granja, R. and M. Wollni. 2018. Dynamics of Smallholder Participation in Horticultural Export Chains: Evidence from Ecuador. *Agricultural economics 49*(2): 225-235.

Guo, L., H. Li, X. Cao, A. Cao, and M. Huang. 2021. Effect of Agricultural Subsidies on the Use of Chemical Fertilizer. *Journal of Environmental Management 299(2)*: 113621.

Gwynne, R. 2008. UK Retail Concentration, Chilean Wine Producers and Value Chains. *Geographical Journal 174*(2): 97-108.

Gyau, A. and A. Spiller. 2008. The Impact of Supply Chain Governance Structures on the Inter-Firm Relationship Performance in Agribusiness. *Zemedelska Ekonomika-Praha 54*(4): 176-185.

Hall, C., B. Knight, S. Ringrose, and O. Knox. 2013. What Have Been the Farm-Level Economic Impacts of the Global Cultivation of GM Crops? *Collaboration for Environmental Science Review* 11-002. www.environmentalevidence.org/SR11002.html. (Accessed March 03, 2023).

Han, X., P. Xue, and N. Zhang. 2021. Impact of Grain Subsidy Reform on the Land Use of Smallholder Farms: Evidence from Huang-Huai-Hai Plain in China. *Land 10*(9): 929.

Hemming, D., E. Chirwa, A. Dorward, H. Ruffhead, R. Hill, J. Osborn, L. Langer, L. Harman, H. Asaoka, and C. Coffey. 2018. Agricultural Input Subsidies for Improving Productivity, Farm Income, Consumer Welfare and Wider Growth in Low-and Lower-Middle-Income Countries: A Systematic Review. *Campbell Systematic Reviews 14*(1): 1-153.

Hou, M., C. Grazia, and G. Malorgio. 2015. Food Safety Standards and International Supply Chain Organization: A Case Study of the Moroccan Fruit and Vegetable Exports. *Food Control 55*: 190-199.

Howeler, R. 2007. Farmer Participation in Research and Extension: The Key to Achieving Adoption of More Sustainable Cassava Production Practices on Sloping Land in Asia and Their Impact on Farmers´ Income. CRC Press. ISBN: 9780429187865.

Hu, X., K. Su, W. Chen, S. Yao, and L. Zhang. 2021. Examining the Impact of Land Consolidation Titling Policy on Farmers' Fertilizer Use: Evidence from a Quasi-Natural Experiment in China. *Land Use Policy 109*: 105645.

Hu, Y., S. Lu, H. Zhang, G. Liu, and J. Peng. 2021. Empirical Analysis on the Performance of Rural Credit Cooperative's Shareholding Reform Based on the Rationale of Isomorphic Incentive Compatibility. *Sustainability 13*(5): 2844.

Hulke, C. and J. Diez. 2022. Understanding Regional Value Chain Evolution in Peripheral Areas Through Governance Interactions–An Institutional Layering Approach. *Applied Geography 139*: 102640.

Ibáñez, L. 2015. Beyond Inclusiveness: Institutions, Cooperation and Rural Development. *Canadian Journal of Development Studies 36*(4): 499-515.

Issahaku, G. and A. Abdulai. 2020. Sustainable Land Management Practices and Technical and Environmental Efficiency Among Smallholder Farmers in Ghana. *Journal of Agricultural and Applied Economics 52*(1): 96-116.

Ito, J. 2015. Diversification of Agricultural Production in China: Economic Rationality of Crop Choice Under the Producer Subsidy Program. *The Japanese Journal of Rural Economics 17*: 1-17.

Jaroensathapornkul, J. 2017. Effects of Non-Tariff Barriers on Thai Exports of Key Vegetables to ASEAN Countries. *International Journal of Trade and Global Markets 10*(1): 3-12.

Jayne, T., T. Yamano, and J. Nyoro. 2004. Interlinked Credit and Farm Intensification: Evidence from Kenya. *Agricultural Economics 31*(2-3): 209-218.

Kaliba, A. and K. Mazvimavi. 2021. Impact of Adopting Improved Seeds on Consumption Expenditure, Income, and Wealth Index Among Small Scale Sorghum Producers in Tanzania. *The Journal of Developing Areas 55*(4): 269-296.

Kangile, J., R. Kadigi, C. Mgeni, B. Munishi, J. Kashaigili, and P. Munishi. 2021. Dynamics of Coffee Certifications in Producer Countries: Re-Examining the Tanzanian Status, Challenges and Impacts on Livelihoods and Environmental Conservation. *Agriculture 11*(10): 931.

Kankwamba, H., M. Kadzamira, and K. Pauw. 2018. How Diversified is Cropping in Malawi? Patterns, Determinants and Policy Implications. *Food Security 10*(2): 323-338.

Kassie, G., B. Rischkowsky, A. Haile, M. Yitayih, G. Legessa, and Y. Fantaye. 2017. Measuring Impact of Access to Livestock Services on Technical Efficiency of Small-Ruminant Production in Rural Ethiopia. *International Livestock Research Institute (ILRI).* https://www.cabdirect.org/cabdirect/abstract/20183342086. (Accessed March 16, 2023).

Khondker, M., H. Ali, V. Upraity, S. Gurung, G. Dhar, and B. Belton. 2018. Making Sense of the Market: Assessing the Participatory Market Chain Approach to Aquaculture Value Chain Development in Nepal and Bangladesh. *Aquaculture 493*: 395-405.

Kinati, W., A. Bekele, and K. Chinnan. 2014. Impact of Farmer Research Group Interventions on Maize Farmers in Central Rift Valley of Oromia: An Empirical Study. *Journal of Agricultural Extension and Rural Development* 6(3): 94-107.

Kinuthia, B. 2020. Agricultural Input Subsidy and Outcomes for Farmers in Tanzania. *WIDER Working Paper* 2020/149. Helsinki. https://doi.org/10.35188/UNU-WIDER/2020/906-8. (Accessed February 20, 2023).

Kirsten, J., M. Mapila, J. Okello, and S. De. 2013. Managing Agricultural Commercialization for Inclusive Growth in Sub-Saharan Africa. *University of Pretoria, Department of Agricultural Economics, Extension and Rural Development.* Working Paper 206518. DOI: 10.22004/ag.econ.206518. (Accessed February 2, 2023).

Krause, H., R. Lippe, and U. Grote. 2016. Adoption and Income Effects of Public GAP Standards: Evidence from the Horticultural Sector in Thailand. *Horticulturae 2*(4): 18.

Kuijpers, R. 2020. Integrated Value Chain Development: Evidence from Bangladesh. *Food Policy 97*: 101916.

Larsen, M. 2016. Sustaining Upgrading in Agricultural Value Chains? State-Led Value Chain Interventions and Emerging Bifurcation of the South Indian Smallholder Tea Sector. *Sustainability* 8(11): 1102.

Lawry, S., C. Samii, R. Hall, A. Leopold, D. Hornby, and F. Mtero. 2014. The Impact of Land Property Rights Interventions on Investment and Agricultural Productivity in Developing Countries: A Systematic Review. *Campbell Systematic Reviews* 10(1): 1-104.

Liang, Q., Y. Xu, X. Wang, S. Jin. 2021. A Driver or a Placebo? The Role of Financial Support in Farmer Cooperative Development in China. *China Agricultural Economic Review* 13(1): 173-192.

Loevinsohn, M., J. Sumberg, A. Diagne, and S. Whitfield. 2013. Under What Circumstances and Conditions Does Adoption of Technology Result in Increased Agricultural Productivity? A Systematic Review. *EPPI-Centre, Social Science Research Unit Institute of Education University of London*. DOI: 10.13140/RG.2.1.3336.4242. (Accessed June 2, 2023).

Lu H., J. Trienekens, S. Omta, and S. Feng. 2010. Guanxi and Contracts in Chinese Vegetable Supply Chains: An Empirical Investigation. *Agri-food chain relationships.* Wallingford UK: CABI: 191-205.

Maggio, G., N. Sitko, and A. Ignaciuk. 2019. Smallholder Adaptive Responses to Seasonal Weather Forecasts. A Case Study of the 2015/16 El Niño Southern Oscillation in Zambia. *FAO Agricultural Development Economics* Working Paper 19-07.

https://www.fao.org/agrifood-economics/publications/detail/en/c/1249380/. (Accessed June 2, 2023).

Maia, A., G. Eusébio, and R. da Silveira. 2020. Can Credit Help Small Family Farming? Evidence from Brazil. *Agricultural Finance Review 80*(2): 212-230.

Maitra, P., S. Mitra, D. Mookherjee, A. Motta, and S. Visaria. 2017. Financing Smallholder Agriculture: An Experiment with Agent-Intermediated Microloans in India. *Journal of Development Economics 127*: 306-337.

Majeed, F., S. Tokgoz, S. Allen, and B. Paris. 2018. Measuring Distortions Along Tanzanian Agricultural Value Chains. *IFPRI-Discussion Papers* (1748). https://www.ifpri.org/publication/measuring-distortions-along-tanzanian-agricultural-value-chains. (Accessed June 15, 2023).

Malan, M., E. Berkhout, J. Duchoslav, M., Voors, and S. van der Esch. 2022. Socioeconomic Impacts of Land Restoration in Agriculture: A Systematic Review. *Ruhr Economic Papers*, no.951. Leibniz-Institut für Wirtschaftsforschung, Essen. ISBN 978-3-96973-114-7.

Mamba, T., C. Machethe, and N. Mtimet. 2018. The Role of Contracts in Improving Access to Credit in the Smallholder Livestock Sector of Swaziland. *International Livestock Research Institute (ILRI).* 37-45. https://cgspace.cgiar.org/handle/10568/95854. (Accessed June 15, 2023).

Manjunatha, A., S., Speelman, S. Aravindakshan, T. Amjath Babu, and P. Mal. 2016. Impact of Informal Groundwater Markets on Efficiency of Irrigated Farms in India: A Bootstrap Data Envelopment Analysis Approach. *Irrigation science* 34: 41-52.

Marr, A., A. Winkel, M. van Asseldonk, R. Lensink, and E. Bulte. 2016. Adoption and Impact of Index-Insurance and Credit for Smallholder Farmers in Developing Countries: A Systematic Review. *Agricultural Finance Review 76(1): 94-118.*

Mason, N., A. Wineman, L. Kirimi, and D. Mather. 2017. The Effects of Kenya's 'Smarter' Input Subsidy Programme on Smallholder Behaviour and Incomes: Do Different Quasi-Experimental Approaches Lead to the Same Conclusions? *Journal of Agricultural Economics 68*(1): 45-69.

Mason, N., A. Wineman, and S. Tembo. 2020. Reducing Poverty by Ignoring the Experts'? Evidence on Input Subsidies in Zambia. *Food Security 12*: 1157-1172.

Matchaya, G. and V. Perotin. 2013. The Impact of Cooperative Patronage: The Case of National Smallholder Farmers' Association (NASFAM) of Malawi in Kasungu District. *Agrekon 52*(2): 75-103.

Mdoda, L., S. Meleni, N. Mujuru, and K. Alaka. 2019. Agricultural Credit Effects on Smallholder Crop Farmers Input Utilization in the Eastern Cape Province, South Africa. *Journal of Human Ecology 66*(1/3): 45-55.

Melese, A. and A. Helmsing. 2010. Endogenisation or Enclave Formation? The Development of the Ethiopian Cut Flower Industry. *The Journal of Modern African Studies 48*(1): 35-66.

Merrey, D. and N. Lefore. 2018. Improving the Availability and Effectiveness of Rural and "Micro" Finance for Small-Scale Irrigation in Sub-Saharan Africa: A Review of Lessons Learned. Colombo, Sri Lanka: *International Water Management Institute* Working Paper 185. doi: 10.5337/2018.225.

Mghenyi, E. 2015. The Impact of Agricultural Credit on Demand for Factors of Production, Farm Output, and Profitability in Kenya. *Thesis, Michigan State University*. https://doi.org/doi:10.25335/M5MD96. (Accessed June 15, 2023).

Mi, Q., X. Li, X. Li, G. Yu, and J. Gao. 2021. Cotton Farmers' Adaptation to Arid Climates: Waiting Times to Adopt Water-Saving Technology. *Agricultural Water Management 244*: 106596.

Minten, B. 2007. Domestic Traders Versus Global Retail Chains: Evidence from Madagascar. *FAO Commodities and Trade Proceedings.* https://agris.fao.org/agris-search/search.do?recordID=XF2016045618. (Accessed June 6, 2023).

Minten, B., A. Taffesse, and P. Brown. 2018. The Economics of Teff: Exploring Ethiopia's Biggest Cash Crop. *International Food Policy Research Institute.* DOI : https://doi.org/10.2499/9780896292833. (Accessed May 12, 2023).

Minviel, J. and L. Latruffe. 2014. Meta-Rregression Analysis of the Impact of Agricultural Subsidies on Farm Technical Efficiency. *European Association of Agricultural Economists Congress*, August 26-29 in Ljubljana, Slovenia. https://ideas.repec.org/p/ags/eaae14/182767.html. (Accessed January 26, 2023).

Mishra, A., S. Shaik, A. Khanal, and S. Bairagi. 2018. Contract Farming and Technical Efficiency: Evidence from Low-Value and High-Value Crops in Nepal. *Agribusiness 34*(2): 426-440.

Missiame, A., R. Nyikal, and P. Irungu. 2021. What is the Impact of Rural Bank Credit Access on the Technical Efficiency of Smallholder Cassava Farmers in Ghana? An Endogenous Switching Regression Analysis. *Heliyon 7*(5): e07102.

Moahid, M., G. Khan, Y. Yoshida, N. Joshi, and K. Maharjan. 2021. Agricultural Credit and Extension Services: Does Their Synergy Augment Farmers' Economic Outcomes? *Sustainability 13*(7): 3758.

Moss, C., A. Lyambabaje, and J. Oehmke. 2017. An Economic Evaluation of SPREAD on Rwanda's Rural Population. *Applied Economics 49*(36): 3634-3644.

Mulwa, C., M. Muyanga, and M. Visser. 2021. The Role of Large Traders in Driving Sustainable Agricultural Intensification in Smallholder Farms: Evidence from Kenya. *Agricultural Economics 52*(2): 329-341.

Muyombano, E. and M. Espling. 2020. Land Use Consolidation in Rwanda: The Experiences of Small-Scale Farmers in Musanze District, Northern Province. *Land Use Policy 99*: 105060.

Nan, Y., Y. Gao, and Q. Zhou. 2019. Rural Credit Cooperatives' Contribution to Agricultural Growth: Evidence from China. *Agricultural Finance Review 79*(1): 119-135.

Nankhuni, F and G. Paniagua. 2013. Meta-Evaluation of Private Sector Interventions in Agribusiness: Finding out What Worked in Access to Finance and Farmer/Business Training. World Bank Working Paper Washington, DC. https://catalog.ihsn.org/index.php/citations/66291. (Accessed April 6, 2023).

Navarra, C. 2017. Contracts Between Smallholders and Private Firms in Mozambique and Their Implications on Food Security. *WIDER Working Paper* 2017/197. Helsinki. https://doi.org/10.35188/UNU-WIDER/2017/423-0. (Accessed December 8, 2022).

Ndegwa, M., A. Shee, C. Turvey, and L. You. 2020. Uptake of Insurance-Embedded Credit in Presence of Credit Rationing: Evidence from a Randomized Controlled Trial in Kenya. *Agricultural Finance* Review 80(5): 745-766.

Ng'asike, P., F. Stepputat, and J. Njoka. 2020. Livestock Trade and Devolution in the Somali-Kenya Transboundary Corridor. *Pastoralism* 10: 1-14.

Ngoma, H., H. Machina, and A. Kuteya. 2021. Can Agricultural Subsidies Reduce Gendered Productivity Gaps? Panel Data Evidence from Zambia. *Development Policy Review 39*(2): 303-323.

Nguyen, A., J. Dzator, and A. Nadolny. 2018. Contract Farming, Agriculture Productivity and Poverty Reduction: Evidence from Tea Estates in Viet Nam. Asia-Pacific Sustainable Development 25(1): 109-145.

Nguyen, H., H. Do, and T. Kompas. 2021. Economic Efficiency Versus Social Equity: The Productivity Challenge for Rice Production in a 'Greying' Rural Vietnam. *World Development 148*: 105658.

Nordjo, R. and C. Adjasi. 2020. The Impact of Credit on Productivity of Smallholder Farmers in Ghana. *Agricultural Finance Review 80*(1): 91-109.

Oberholster, C., C. Adendorff, and K. Jonker. 2015. Financing Agricultural Production from a Value Chain Perspective: Recent Evidence from South Africa. *Outlook on Agriculture 44*(1): 49-60.

Ogundeji, A., E. Donkor, C. Motsoari, and S. Onakuse. 2018. Impact of Access to Credit on Farm Income: Policy Implications for Rural Agricultural Development in Lesotho. *Agrekon 57*(2): 152-166.

Okry, F., P. Van Mele, E. Nuijten, P. Struik, and R. Mongbo. 2011. Organizational Analysis of the Seed Sector of Rice in Guinea: Stakeholders, Perception and Institutional Linkages. *Experimental Agriculture 47*(1): 137-157.

Onyango, C., J. Nyaga, J. Wetterlind, M. Söderström, and K. Piikki. 2021. Precision Agriculture for Resource use Efficiency in Smallholder Farming Systems in Sub-Saharan Africa: A Systematic Review. *Sustainability 13*(3): 1158.

Orimogunje, R., A. Ogunleye, and A. Kehinde. 2020. Effect of Microcredit on Profit Efficiency of Small-Scale Poultry Farmers Oyo State, Nigeria. *Agricultura 17*(1-2): 37-46.

Ortiz, O., R. Orrego, W. Pradel, P. Gildemacher, R. Castillo, R. Otiniano, J. Gabriel, J. Vallejo, O. Torres, and G. Woldegiorgis. 2013. Insights into Potato Innovation Systems in Bolivia, Ethiopia, Peru and Uganda. *Agricultural Systems 114*: 73-83.

Ouma, E., J. Ochieng, M. Dione, and D. Pezo. 2017. Governance Structures in Smallholder Pig Value Chains in Uganda: Constraints and Opportunities for Upgrading. *International Food and Agribusiness Management Revie 20*(3): 307-319.

Owusu, V. and A. Abdulai. 2019. Examining the Economic Impacts of Integrated Pest Management Among Vegetable Farmers in Southern Ghana. *Journal of Environmental Planning and Management 62*(11): 1886-1907.

Oya, C., F. Schaefer, and D. Skalidou. 2018. The Effectiveness of Agricultural Certification in Developing Countries: A Systematic Review. *World Development* 112: 282-312.

Padron, B. 2012. Institutions in the Mexican Coffee Sector: Changes and Responses. Thesis, Wageningen University and Research. https://edepot.wur.nl/238636. (Accessed August, 17, 2022).

Pande, D. 2019. Impact of the Agriculture Debt Waiver Debt Relief Scheme on the Consumption Expenditure of Smallholder Farmers. *Agricultural Economics Research Review*, *32*(conf): 103-112.

Pandey, V., A. Gupta, and S. Gupta. 2019. Labor and Welfare Impacts of a Large-Scale Livelihoods Program: Quasi-Experimental Evidence from India. *World Bank Policy Research* Working Paper 8883. https://doi.org/10.1596/1813-9450-8883. (Accessed December 23, 2022).

Pavez, I., J. Codron, P. Lubello, and M. Florêncio. 2019. Biosecurity Institutions and the Choice of Contracts in International Fruit Supply Chains. *Agricultural Systems 176*: 102668.

Peprah, J., I. Koomson, J. Sebu, and C. Bukari. 2020. Improving Productivity Among Smallholder Farmers in Ghana: Does Financial Inclusion Matter? *Agricultural Finance Review 81*(4): 481-502.

Perera, S., S. Dissanayake, D. Rajapaksa, and A. Lankapura. 2021. Does Microcredit Play a Role in Improving the Technical Efficiency of Paddy Farmers? A Case Study in the Anuradhapura District of Sri Lanka. *Sri Lankan Journal of Agriculture and Ecosystems* 3(1): 167–182.

Posthumus, H., A. Martin, and T. Chancellor. 2013. A Systematic Review on the Impacts of Capacity Strengthening of Agricultural Research Systems for Development and the

Conditions of Success. Evidence for Policy and Practice Information and Coordinating Centre (EPPI-Centre), University of London. http://gala.gre.ac.uk/id/eprint/9689. (Accessed December 2, 2022).

Prifti, E., S. Daidone, N. Pace, and B. Davis. 2020. Stuck Exchange: Can Cash Transfers Push Smallholders Out of Autarky? *The Journal of International Trade & Economic Development* 29(5): 495-509.

Quarmine, W. 2013. Incentives For Smallholders to Enhance the Production of Quality Cocoa Beans in Ghana: The Role of Institutions. Thesis, Wageningen University. https://core.ac.uk/download/pdf/29217306.pdf. (Accessed March 12, 2023).

Ragasa, C., I. Lambrecht, and D. Kufoalor. 2018. Limitations of Contract Farming as a Pro-Poor Strategy: The Case of Maize Outgrower Schemes in Upper West Ghana. *World Development* 102: 30-56.

Rao, N., R. Sutradhar, and T. Reardon. 2017. Disruptive Innovations in Food Value Chains and Small Farmers in India. Indian Journal of Agricultural Economics 72(1): 24-48.

Rashid, S., N. Tefera, N. Minot, N., and G. Ayele. 2013. Can Modern Input Use Be Promoted Without Subsidies? An Analysis of Fertilizer in Ethiopia. *Agricultural Economics* 44(6): 595-611.

Rich, K. and B. Bennett. 2019. Using Preferential Trade Access to Promote Global Development Goals: The Case of Beef and Market Access to Norway from Namibia and Botswana. *Agrekon* 58(4): 485-502.

Ricker-Gilbert, J. and T. Jayne. 2017. Estimating the Enduring Effects of Fertilizer Subsidies on Commercial Fertilizer Demand and Maize Production: Panel Data Evidence From Malawi. *Journal of Agricultural Economics* 68(1): 70-97.

Rocha, J. 2017. Agricultural Extension, Technology Adoption and Household Food Security: Evidence from DRC. Thesis – Wageningen University and Research. DOI: 10.18174/416256. (Accessed April 20, 2023).

Rodrigo, M. 2013. Three Essays on Agricultural Policy and Development Economics. Dissertation, The University of Wisconsin-Madison. https://www.proquest.com/docview/1433929796?pq-origsite=gscholar&fromopenview=true. (Accessed January 23, 2023).

Saenger, C., M. Torero, and M. Qaim. 2012. Impact of Third-Party Enforcement of Contracts in Agricultural Markets–a Field Experiment in Vietnam. *American Journal for Agricultural Economics* 96(4): 1220-1238.

Sakprachawut, S. and D. Jourdain. 2016. Land Titles and Formal Credit in Thailand. *Agricultural Finance Review* 76(2): 270-287.

Sam, V. 2021. Formal Credit Usage and Gender Income Gap: The Case of Farmers in Cambodia. *Agricultural Finance Review* 81(5): 675-701.

Schuster, M. and M. Maertens. 2016. Do Private Standards Benefit Workers in Horticultural Export Chains in Peru? *Journal of Cleaner Production 112(4)*: 2392-2406.

Sekhri, S. 2022. Agricultural Trade and Depletion of Groundwater. *Journal of Development Economics*, *156*: 102800. https://doi.org/https://doi.org/10.1016/j.jdeveco.2021.102800. (Accessed November 29, 2022).

Seleka, T. 2022. Old Wine in a New Bottle? Impact of the ISPAAD Input Subsidy Program on the Subsistence Economy in Botswana. Review of Development Economics 26(1): 298-322.

Shashi, K. and M. Vigneri. 2017. The Cocoa Coast: The Board-Managed Cocoa Sector in Ghana. International Food Policy Research Institute (IFPRI). DOI: https://doi.org/10.2499/9780896292680. (Accessed May 29, 2023).

Sheahan, M., J. Ariga, and T. Jayne. 2016. The Rewards of an Improved Enabling Environment: How Input Market Reform Helped Kenyan Farmers Raise Their Fertilizer Use By 36%. *Feed the Future Innovation Lab for Food Security Policy Research Briefs 260042*, Michigan State University. DOI: 10.22004/ag.econ.260042. (Accessed May 12, 2023).

Sher, A., S. Mazhar, H. Azadi, and G. Lin. 2020. Smallholder Commercialization and Urban-Rural Linkages: Effect of Interest-Free Agriculture Credit on Market Participation of Rice Growers in Pakistan. *Land* 10(1): 1-24.

Sianjase, A., and V. Seshamani. 2013. Impacts of Farmer Inputs Support Program on Beneficiaries in Gwembe District of Zambia. *Journal of Environmental Issues and Agriculture in Developing Countries* 5(1): 40-50.

Sibande, L., A. Bailey, and S. Davidova. 2017. The Impact of Farm Input Subsidies on Maize Marketing in Malawi. *Food Policy* 69: 190-206.

Silva, A., A. Barbosa, and C. Fontes. 2014. Certification Rules for the Fruit Agri-Business. *African Journal of Agricultural Research* 9(37): 2805-2813.

Sinyolo, S., E. Wale, and M. Mudhara. 2016. The Impact of Social Grants on the Propensity and Level of Use of Inorganic Fertilizer Among Smallholders in Kwazulu-Natal, South Africa. *Agrekon* 55(4): 436-457.

Sitko, N., A. Scognamillo, and G. Malevolti. 2021. Does Receiving Food Aid Influence the Adoption of Climate-Adaptive Agricultural Practices? Evidence from Ethiopia and Malawi. *Food Policy* 102: 102041.

Songsermsawas, T. 2015. Factors Affecting Farm Productivity in Rural India: Social Networks and Market Access. Dissertation, University of Illinois at Urbana-Champaign. https://hdl.handle.net/2142/88037. (Accessed April 2, 2023).

Soullier, G. and P. Moustier. 2018. Impacts of Contract Farming in Domestic Grain Chains on Farmer Income and Food Insecurity. Contrasted Evidence from Senegal. *Food Policy* 79: 179-198.

Stewart, R., L. Langer, N. Da Silva, E. Muchiri, H. Zaranyika, Y. Erasmus, N. Randall, S. Rafferty, M. Korth, N. Madinga, and T. de Wet. 2015. The Effects of Training, Innovation and New Technology on African Smallholder Farmers' Economic Outcomes and Food Security: A Systematic Review. *Campbell Systematic Reviews* 11(1): 1-224. https://doi.org/https://doi.org/10.4073/csr.2015.16. (Accessed May 2, 2023).

Takeshima, H. and A. Kumar. 2020. Changing Returns-to-Scale and Deepening of Factor-Endowments-Induced Specialization: Exploring Broader Linkage Between Agricultural Mechanization and Agricultural Transformation in Nepal. International Food Policy Research Institute. (Vol. 1934). DOI: https://doi.org/10.2499/p15738coll2.133757. (Accessed April 1, 2023).

Takeshima, H. and L. Liverpool-Tasie. 2015. Fertilizer Subsidies, Political Influence and Local Food Prices in Sub-Saharan Africa: Evidence from Nigeria. *Food Policy* 54: 11-24.

Tamini, L., I. Bocoum, G. Auger, K. Lawin, and A. Traoré. 2019. Enhanced Microfinance Services and Agricultural Best Management Practices: What Benefits for Smallholders Farmers? An Evidence from Burkina Faso. CIRANO Working Papers-Center for Interuniversity Research and Analysis on Organizations (2019s-11). https://cirano.qc.ca/files/publications/2019s-11.pdf. (Accessed December 17, 2022).

Teklewold, H. 2021. How Effective is Ethiopia's Agricultural Growth Program at Improving the Total Factor Productivity of Smallholder Farmers? *Food Security* 13(4): 895-912.

Tesfay, M. 2021. The Impact of Participation in Rural Credit Program on Adoption of Inorganic Fertilizer: A Panel Data Evidence from Northern Ethiopia. *Cogent Food & Agriculture* 7(1): 1919388.

Theriault, V. and M. Smale. 2021. The Unintended Consequences of the Fertilizer Subsidy Program on Crop Species Diversity in Mali. *Food Policy* 102: 102121.

Thériault, V., M. Smale, and A. Assima. 2018. The Malian Fertilizer Value Chain Post-Subsidy: An Analysis of its Structure and Performance. *Development in Practice* 28(2): 242-256.

Ton, G., K. de Grip, L. Klerkx, M. Rau, M. Douma, E. Friis-Hansen, B. Triomphe, A. Waters-Bayer, and M. Wongtschowski. 2013. Effectiveness of Innovation Grants to Smallholder Agricultural Producers: An Explorative Systematic Review. EPPI Centre, Social Science Research Unit, Institute of Education, University of London. ISBN: 978-1-907345-54-8.

Ton, G., S. Desiere, W. Vellema, S. Weituschat, and M. D'Haese. 2017. The Effectiveness of Contract Farming for Raising Income of Smallholder Farmers in Low- and Middle-Income Countries: A Systematic Review. *Campbell Systematic Reviews* 13(1): 1-131. https://doi.org/https://doi.org/10.4073/csr.2017.13. (Accessed June 12, 2023).

Totin, E., B. van Mierlo, R. Mongbo, and C. Leeuwis. 2015. Diversity in Success: Interaction Between External Interventions and Local Actions in Three Rice Farming Areas in Benin. *Agricultural Systems* 133: 119-130.

Tran, N., C. Bailey, N. Wilson, and M. Phillips. 2013. Governance of Global Value Chains in Response to Food Safety and Certification Standards: The Case of Shrimp from Vietnam. *World Development* 45: 325-336.

Tsiboe, F., I. Egyir, and G. Anaman. 2021. Effect of Fertilizer Subsidy on Household Level Cereal Production in Ghana. *Scientific African* 13: e00916.

Utting-Chamorro, K. 2005. Does Fair Trade Make a Difference? The Case of Small Coffee Producers in Nicaragua. *Development in Practice* 15(3-4): 584-599.

Van Campenhout, B., B. Minten, and J. Swinnen. 2021. Leading the Way–Foreign Direct Investment and Dairy Value Chain Upgrading in Uganda. *Agricultural Economics* 52(4): 607-631.

Van Ewijk, E. and M. Ros-Tonen. 2021. The Fruits of Knowledge Co-Creation in Agriculture and Food-Related Multi-Stakeholder Platforms in Sub-Saharan Africa–A Systematic Literature Review. *Agricultural Systems* 186: 102949.

Verdier-Chouchane, A. and A. Boly. 2017. Feeding Africa: Towards Agro-Allied Industrialization for Inclusive Growth. *African Development Review* 29(s2): 75-212.

Vilas-Boas, J., L. Klerkx, R. Lie. 2022. Connecting Science, Policy, and Practice in Agri-Food System Transformation: The Role of Boundary Infrastructures in the Evolution of Brazilian Pig Production. *Journal of Rural Studies 89*: 171-185.

Waddington, H., B. Snilstveit, J. Hombrados, M. Vojtkova, D. Phillips, P. Davies, and H. White. 2014. Farmer Field Schools for Improving Farming Practices and Farmer Outcomes: A Systematic Review. *Campbell Systematic Reviews 10*(1): i-335.

Wahyudin, M., M. Maksum, and H. Yuliando. 2015. The Shallot Pricing in the View of Import Restriction and Price Reference. *Agriculture and Agricultural Science Procedia 3*: 132-136.

Walls, H., D. Johnston, M. Tak, J. Dixon, J. Hanefeld, E. Hull, and R. Smith. 2018. The Impact of Agricultural Input Subsidies on Food and Nutrition Security: a Systematic Review. *Food Security* 10(6): 1425-1436. https://doi.org/10.1007/s12571-018-0857-5. (Accessed June 1, 2023).

Wang, J., A. Siraj, F. Shah, and T. Huang. 2019. Economic Impact of Formal Credit Constraints on Technical Efficiency of Wheat: Evidence from Sindh Pakistan. *Fresenius Environmental Bulletin* 28(8): 5888-5899.

Wang, Y., R. Bilsborrow, Q. Zhang, J. Li, and C. Song. 2019. Effects of Payment for Ecosystem Services and Agricultural Subsidy Programs on Rural Household Land Use Decisions in China: Synergy or Trade-Off? *Land Use Policy* 81: 785-801.

Weatherspoon, D., S. Miller, F. Niyitanga, L. Weatherspoon, and J. Oehmke. 2021. Rwanda's Commercialization of Smallholder Agriculture: Implications for Rural Food Production and Household Food Choices. *Journal of Agricultural & Food Industrial Organization* 19(1): 51-62.

Weyori, A., M. Amare, H. Garming, and H. Waibel. 2018. Agricultural Innovation Systems and Farm Technology Adoption: Findings from a Study of the Ghanaian Plantain Sector. *The Journal of Agricultural Education and Extension* 24(1): 65-87.

Wong, H., X. Wei, H. Kahsay, Z. Gebreegziabher, C. Gardebroek, D. Osgood, and R. Diro. 2020. Effects of Input Vouchers and Rainfall Insurance on Agricultural Production and Household Welfare: Experimental Evidence from Northern Ethiopia. *World Development* 135: 105074.

Wordofa, M., J. Hassen, G. Endris, C. Aweke, D. Moges, and D. Rorisa. 2021. Adoption of Improved Agricultural Technology and Its Impact on Household Income: A Propensity Score Matching Estimation in Eastern Ethiopia. *Agriculture & Food Security* 10: 1-12.

Wordofa, M., J. Hassen, G. Endris, C. Aweke, D. Moges, and D. Rorisa. 2021. Impact of Improved Agricultural Technology Use on Household Income in Eastern Ethiopia: Empirical Evidence from a Propensity Score Matching Estimation. *Journal of Land and Rural Studies* 9(2): 276-290.

Wossen, T., T. Abdoulaye, A. Alene, M. Haile, S. Feleke, A. Olanrewaju, and V. Manyong. 2017. Impacts of Extension Access and Cooperative Membership on Technology Adoption and Household Welfare. *Journal of Rural Studies* 54: 223-233.

Yadav, S., K. Sharma, J. Yadav, and V. Kumar. 2022. Impact of National Horticulture Mission on the Status of Beneficiary Farmers in Rajasthan District of India. Indian Research *Journal of Extension Education* 22(1): 134-137.

Yaseen, A., K. Bryceson, and A. Mungai. 2018. Commercialization Behaviour in Production Agriculture: The Overlooked Role of Market Orientation. *Journal of Agribusiness in Developing and Emerging Economies* 8(3): 579-602.

Yi, F., D. Sun, and Y. Zhou. 2015. Grain Subsidy, Liquidity Constraints and Food Security—Impact of the Grain Subsidy Program on the Grain-Sown Areas in China. *Food Policy* 50: 114-124.

Zeng, Y., F. Jia, L. Wan, and H. Guo. 2017. E-Commerce in Agri-Food Sector: A Systematic Literature Review. *International Food and Agribusiness Management Review* 20(4): 439-460.

Zhang, Q., B. Yan, and X. Huo. 2018. What are the Effects of Participation in Production Outsourcing? Evidence from Chinese Apple Farmers. *Sustainability* 10(12): 4525.

Zhou, L., L. Li, and L. Lei. 2019. Avian Influenza, Non-Tariff Measures and the Poultry Exports of China. *Australian Journal of Agricultural and Resource Economics* 63(1): 72-94.

Zhu, Q., C. Wachenheim, Z. Ma, Z., and C. Zhu. 2018. Supply Chain Re-Engineering: A Case Study of the Tonghui Agricultural Cooperative in Inner Mongolia. *International Food and Agribusiness Management Review*21(1): 133-160.

Appendix B3. Questionnaire for Inclusion of Studies

Section 1 of 8.

This form was designed to collect data from the included studies in the EGM on agricultural policies and agricultural value chains to address Question 3 in the USDA Food for Progress Learning Agenda.

Research question 1: What are the extent and characteristics of existing empirical evidence regarding the policies that enhance AVC and improve enabling environments and what are the evidence gaps in the literature?

Research question 2: What are the extent and characteristics of existing empirical evidence regarding the effectiveness of models of collaboration among local and international actors, including donors, private sector partners, academic institutions, and NGOs to support policy change?

Section 2 of 8

Identification.

2.1. Reviewer.

2.2. Covidence ID number (number only).

2.3. Name of the Journal/Publisher.

2.4. Type of publication.

1 = Journal

2 = Unpublished report

3 = Dissertation/thesis

4 = Book/chapter

5 =Meeting presentation

99 = Other

2.5. Year of publication (YYYY).

2.6. Authors' last name (Examples: 1. For two authors Abram and Albuja; 2. For more than 2 authors Di et al.).

2.7. Study title.

2.8. Objective of the study.

2.9. Country.

2.10. Does the study provide the evidence on

1= EGM main question (the effect of policy) only (go to section 5 - evidence on question 1).

2= EGM supplementary question (coordination) only (go to section 4 - evidence on question 2)

3= Both questions (go to section 4 - evidence on question 2)

4= The study is irrelevant (go to section 3 – reasons for exclusion)

Section 3 of 8 Reasons for exclusion of irrelevant study.

3.1. Explain why this study should be excluded.

Section 4 of 8 Evidence on research question 2. What are the extent and characteristics of existing empirical evidence regarding the effectiveness of models of collaboration among local and international actors, including donors, private sector partners, academic institutions, and NGOs to support policy change?

4.1. Brief description of the policy or policy change process.

4.2. What are the bodies that worked towards value chain improvement policies? (Check all that apply).

1= Individual farmers
2= Farmer organizations/cooperatives

3= Processors
4= Traders
5= Supermarkets (local or nationals)
6= Consumers
7= industry partners
8= Local donors
9= Local governments
10=Local research or education institutions
11= National institutions
12= International NGOs or International Development Agency
13= Unclear

4.3. Brief description of the collaboration model presented in the study.

4.4. In what stage, did collaboration become crucial?
1= Policy design and planning
2= Targeting
3= Policy promotion/adoption
4= Resource mobilization
5= Policy implementation and service delivery
6= Policy monitoring
7= Policy evaluation and review
8= Unclear

4.5. What was the overall impact of collaboration model?

1= Positive
2= Negative
3= Neutral
4= Unclear

4.6. From governance perspective, which governance principles do the authors discuss as the core of the collaborative partnership? (Check Bovaird 2004).

1= Citizen engagement
2= Transparency
3= Accountability
4= Equalities and social inclusion
5= Ethical and honest behavior
6= Equity (fair procedures and due process)
7= Willingness and ability to collaborate
8= Ability to compete
9= Leadership

10= Sustainability
11= None

4.7. What stage of the value chain was meant to be improved as a result of collaboration?

1= Input supply
2= Production/Post-harvest
3= Processing/logistics
4= Markets
5= Governance in any tier/link of the value chain

4.8. What are the factors that encourage stakeholders to collaborate (enablers of collaboration)? Examples could include providing a public good/service, pre-existing relationships, need of ending a conflict, specific incentive.

4.9. Which quantitative research methods did the authors use to study the model of collaboration (if any)?

4.10. Which qualitative research methods did the authors use to study the model of collaboration?

4.11. Comments.

4.12. Does the study provide any evidence on the policy effectiveness (Question 1)?

1= Yes (go to section 5)

2= No (submit form)

Section 5 of 8 Policy.

5.1. Name of the policy (if stated, provide the full name and abbreviations in the parenthesis).

5.2. Goal of the evaluated policy or policy change.

1= Policy (or policy change) leadership

2= Governments (national or local)

3= International NGOs or International Development Agency

4= Stakeholders (through agreements like contract farming, certifications, information services)

5= Unclear

5.3. Related agriculture.

1= Crops
2= Livestock (farm animals with the exception of poultry)
3= Poultry
4= Farm fishing and fisheries
5= Not specified

5.4. Specify the name of studied crops or livestock.

5.5. The policy focuses on

1= One of the stakeholders' groups in the value chain (usually the farmers)
2= A broader intervention in the value chain (more than one tier/link or the whole value chain)

5.6. Which broad category best describes the policy scope? (Check all that apply). Please check the protocol AND Dr. Norton's books for details.

1= Policies that influence farmer's incentives (subsidies, prices, market access, information access, etc.)
2= Land tenure policies (land reform, land rights, land markets, access to land, etc.)
3= Policies for management of resources (land, water, forest, fisheries)
4= Policies for agricultural and rural finance and/or risk management (credit, insurance, financial portfolios)
5= Policies for agricultural technology (research, extension, innovation, etc.)
6= Policies that influence processor's/trader's incentives (prices, trade conditions, exchange rate, fiscal policy, etc.)
7= Policies to create/strength value addition services/capacities (grading, packaging, crop processing, etc.)
8= Policies for competitiveness of exports (quality upgrading, standards, quotes administration, SFS policies, distribution network)
9= Supporting policies for competitiveness (infrastructure, transportation, connectivity, logistics)
10=Sustainability of the value chain (environmental concerns, waste management)
11=Governance of the value chain (relative power of stakeholders)
12=Marketing policies and other consumer-oriented policies
13=Other

5.7. Which specific category best describes the policy? (Check all that apply). Please check the protocol for details.

1= Input distribution network and management
2= Information sharing and knowledge management
3= Credit, subsidies, and financial services
4= Land tenure and/or land improvement
5= Expert services in production and post-harvest stages
6= Harvesting or distribution of fruits, vegetables, and other high-value crops
7= Dairy, meat and livestock related services

8= Value addition services (grading, sorting, packing, etc)
9= Transportation and shipment related services
10= Product traceability services
11= Contract farming and industry tie-up
12= Certification along value chain
13= Commodity markets and trading platforms (physical and online)
14= Price support and market information network/channels
15= Intervention related to gender issues, poor households or social groups
16= Governance and decentralization of value chain
17= Trade policies
18=Other

Section 6 of 8

Outcomes.

For this section, please do a careful revision of the results section in the study (especially tables and supplementary material). Mark options for estimated quantitative or discussed qualitative effects, independent of its direction (sign), magnitude or significance.

6.1. Does the study reports effect of the policy on any of the following production outcomes? (Check all that apply. For option "other" please use the following style: (1) name of outcome 1; (2) name of outcome 2; etc.).

1= Yield
2= Technology adoption
3= Area under cultivation
4= Crop diversification
5= It does not report production outcomes

6.2. Does the study reports effect of the policy on any of the following market reform and expansion outcomes? (Check all that apply. For option "other" please use the following style: (1) name of outcome 1; (2) name of outcome 2; etc.). (Check all that apply. For option "other" please use the following style: (1) name of outcome 1; (2) name of outcome 2; etc.).

1= Market information on demand and prices
2= Market access to sell produce
3= Market development, market through farmer groups, new markets
4= E-markets
5= Profit or income from product sales - Agricultural income
6= Share of profit (along the value chain)
7= Cross-border trades
8= It does not report market reform and expansion outcomes

6.3. Does the study reports effect of the policy on any of the following logistic improvement and product improvement outcomes? (Check all that apply. For option "other" please use the following style: (1) name of outcome 1; (2) name of outcome 2; etc.).

1= Logistics for inputs availability
2= Produce procurement services
3= Reduction in food waste
4= Product quality
5= Transportation facility
6= It does not report logistic/product improvement outcomes

6.4. Does the study reports effect of the policy on any of the following risk management outcomes? (Check all that apply. For option "other" please use the following style: (1) name of outcome 1; (2) name of outcome 2; etc.).

1= Yield risk
2= Price risk and uncertainty
3= Exchange rate risk
4= Contract risk
5= Climate change adaptation
6= It does not report risk management outcomes

6.5. Does the study reports effect of the policy on any of the following welfare outcomes? (Check all that apply. For option "other" please use the following style: (1) name of outcome 1; (2) name of outcome 2; etc.).

1= Family income (includes off-farm income and from other sources
2= Food security
3= Nutritional security
4= Food access
5= Education/schooling
6= Poverty reduction
7= It does not report welfare outcomes

6.6. Was any other relevant outcome reported?

Section 7 of 8 Study design.

7.1. Which methods does the study mainly use?

1= Quantitative methods
2= Qualitative methods

3= Mixed methods

7.2. For quantitative studies, name of the (independent) variable that represents the policy.

7.3. For quantitative studies, is the (independent) variable that represents the policy a dummy variable?

1= Yes
2= No

7.4. Focused on the main estimation, the study uses

1= Single method for single outcome
2= Single method for multiple outcomes
3= Multiple methods for single outcome
4= Multiple methods for multiple outcomes

7.5. Which quantitative methods did the authors use? (Choose any combination that applies for the MAIN analytical method to establish the effect of the policy in the ag value chain; Check "Other" and fill out "NA" for exclusively qualitative studies).

1= PSM
2= RCT
3= DID
4= IV
5= RDD
6= Other

7.6. Which qualitative research methods did the authors use to study the effect of the policy? (NA for exclusively quantitative studies).

Section 8 of 8 Comments.

8.1. Comments.

Appendix B4. Questionnaire for Inclusion of Systematic Reviews

Section 1 of 6.

This form was designed to apply inclusion / exclusion criteria on the retrieved Systematic Reviews and collect basic data from the included ones.

Research question 1: What are the extent and characteristics of existing empirical evidence regarding the policies that enhance AVC and improve enabling environments and what are the evidence gaps in the literature?

Research question 2: What are the extent and characteristics of existing empirical evidence regarding the effectiveness of models of collaboration among local and international actors, including donors, private sector partners, academic institutions, and NGOs to support policy change?

Section 2 of 6

Identification.

2.1. Reviewer.

2.2. Study ID number (number only).

2.3. Year of publication (YYYY).

2.4. Authors' last name (Examples: 1. For two authors Abram and Albuja; 2. For more than 2 authors Di et al.).

2.5. SR title.

2.6. Geographical scope.

Section 3 of 6

Decision.

3.1. This SR should be...

1= Included (go to section 4)
2= Excluded (go to section 5)

3= Discusses (go to section 6)

Section 4 of 6 Location on the intervention-outcome matrix.

4.1. Which broad category best describes the interventions considered in the SR? (Check all that apply).

1= Policies that influence farmer's incentives (subsidies, prices, market access, information access, etc.)
2= Land tenure policies (land reform, land rights, land markets, access to land, etc.)
3= Policies for management of resources (land, water, forest, fisheries)
4= Policies for agricultural and rural finance and/or risk management (credit, insurance, financial portfolios)
5= Policies for agricultural technology (research, extension, innovation, etc.)
6= Policies that influence processor's/trader's incentives (prices, trade conditions, exchange rate, fiscal policy, etc.)
7= Policies to create/strength value addition services/capacities (grading, packaging, crop processing, etc.)
8= Policies for competitiveness of exports (quality upgrading, standards, quotes administration, SFS policies, distribution network)
9= Supporting policies for competitiveness (infrastructure, transportation, connectivity, logistics)
10=Sustainability of the value chain (environmental concerns, waste management)
11=Governance of the value chain (relative power of stakeholders)
12=Marketing policies and other consumer-oriented policies
13=Other

4.2. Which specific category best describes the policy? (Check all that apply). Please check the protocol for details.

1= Input distribution network and management
2= Information sharing and knowledge management
3= Credit, subsidies, and financial services
4= Land tenure and/or land improvement
5= Expert services in production and post-harvest stages
6= Harvesting or distribution of fruits, vegetables, and other high-value crops
7= Dairy, meat and livestock related services
8= Value addition services (grading, sorting, packing, etc)
9= Transportation and shipment related services
10= Product traceability services
11= Contract farming and industry tie-up
12= Certification along value chain
13= Commodity markets and trading platforms (physical and online)
14= Price support and market information network/channels
15= Intervention related to gender issues, poor households or social groups
16= Governance and decentralization of value chain
17= Trade policies

18=Other

4.3. Does the study reports effect of the intervention on any of the following outcomes? (Check all that apply. For option "other" please use the following style: (1) name of outcome 1; (2) name of outcome 2; etc.).

1= Production outcomes
2= Market reform and expansion outcomes
3= Logistic improvement and product improvement outcomes
4= Risk management outcomes
5= Welfare outcomes

Section 5 of 6 Exclusion.

5.1. Explain why this SR should be excluded

Section 6 of 6 Comments.

6.1. Comments.

Appendix B5. Questionnaire for Quality Assessment of Systematic Reviews

Section 1 of 6. Identification.

1.1. Reviewer.

1.2. Study ID number (number only).

1.3. Year of publication (YYYY).

1.4. Authors' last name (Examples: 1. For two authors Abram and Albuja; 2. For more than 2 authors Di et al.).

1.5. SR title.

Section 2 of 4 Methods used to identify, include, and critically appraise studies.

2.1. a) Did the authors specify (criteria)
1= Types of studies
2= Participants/ settings/ population
3= Intervention(s)
4= Outcome(s)

2.1. b) Were the criteria used for deciding which studies to include in the review reported?

1= Yes (All four should be yes)
2= No (All four should be no)
3= Partially (Any other)

2.2. a) Were the following done:

1= Language bias avoided (no restriction of inclusion based on language)
2= No restriction of inclusion based on publication status
3= Relevant databases searched (Minimum criteria: All reviews should search at least one source of grey literature such as Google; for health: Medline/ Pubmed + Cochrane Library; for social sociences: IDEAS + at least one database of general social science literature and one subject specific database)
4= Reference lists in included articles checked
5= Authors/experts contacted

2.2. b) Was the search for evidence reasonably comprehensive?

1= Yes (All five should be yes)
2= Partially (Relevant databases and reference lists are both reported)
3= No (Any other)
4= Can't tell

2.3. Does the review cover an appropriate time period?

1= Yes
2= No
3= Can't tell ((only use if no information about time period for search)
4= Unsure

2.4. a) Did the authors specify

1= Independent screening of full text by at least 2 reviewers
2= List of included studies provided
3= List of excluded studies provided

2.4.b) Was bias in the selection of articles avoided?

1= Yes (All three should be yes, although reviews published in journals are unlikely to have a list of excluded studies (due to limits on word count) and the review should not be penalized for this.)
2= No (All other)
3= Partially (Independent screening and list of included studies provided are both reported.)

2.5. a) What criteria were used for the quality assessment

1= The criteria used for assessing the quality/ risk of bias were reported
2= A table or summary of the assessment of each included study for each criterion was reported
3= Sensible criteria were used that focus on the quality/ risk of bias (and not other qualities of the studies, such as precision or applicability/external validity). "Sensible" is defined as a recognized quality appraisal tool/ checklist, or similar tool which assesses bias in included studies.

2.5.b) Did the authors use appropriate criteria to assess the quality and risk of bias in analysing the studies that are included?

1= Yes
2= No
3= Partially

2.6. Comments

Section 3 of 4 Methods used to analyze the findings

3.1. a) Was there

1= Independent data extraction by at least 2 reviewers
2= A table or summary of the characteristics of the participants, interventions and outcomes for the included studies
3= A table or summary of the results of all the included studies

3.1. b) Were the characteristics and results of the included studies reliably reported?

1= Yes (All three should be yes)
2= No (None of these are reported. If the review does not report whether data was independently extracted by 2 reviewers (possibly a reporting error), we downgrade to NO)
3= Partially (Criteria one and three are yes, but some information is lacking on second criteria.)
4= Not applicable (e.g. no included studies)

3.2. Are the methods used by the review authors to analyze the findings of the included studies clear, including methods for calculating effect sizes if applicable?

1= Yes (Methods used clearly reported. If it is clear that the authors use narrative synthesis, they don't need to say this explicitly)
2= No (Nothing reported on methods)
3= Partially (Some reporting on methods but lack of clarity)
4= Not applicable (if no studies/no data)

3.3. a) Did the review considers any of the following questions?

1= Did the review ensure that included studies were similar enough that it made sense to combine them, sensibly divide the included studies into homogeneous groups, or sensibly conclude that it did not make sense to combine or group the included studies?
2= Did the review discuss the extent to which there were important differences in the results of the included studies?
3= If a meta-analysis was done, was the I2, chi square test for heterogeneity or other appropriate statistic reported? If no statistical test was reported, is a qualitative justification made for the use of random effects?

3.3.b) Did the review describe the extent of heterogeneity?

1= Yes (First two should be yes, and third category should be yes if applicable should be yes)
2= No (Any other)
3= Partially (The first category is yes)

4= Not applicable (e.g. no studies or no data)

3.4. How was the data analysis done?

1= Descriptive only
2= Vote counting based on direction of effect
3= Vote counting based on statistical significance
4= Description of range of effect sizes
5= Meta-analysis
6= Meta-regression
7= Not applicable (e.g. no studies or no data)

3.5. How were the studies weighted in the analysis?

1= Equal weights (this is what is done when vote counting is used)
2= By quality or study design (this is rarely done)
3= Inverse variance (this is what is typically done in a meta-analysis)
4= Number of participants (sample size)
5= Not clear
6= Not applicable (e.g. no studies or no data)

3.6. Did the review address unit of analysis errors?

1= Yes - took clustering into account in the analysis (e.g. used intra-cluster correlation coefficient)
2= No, but acknowledged problem of unit of analysis errors
3= No mention of issue
4= Not applicable - no clustered trials or studies included

3.7. Were the findings of the relevant studies combined (or not combined) appropriately relative to the primary question the review addresses and the available data?

1= Yes (if appropriate table, graph or meta analysis and appropriate weights and unit of analysis errors addressed (if appropriate)
2= No (if narrative or vote counting, where quantitative analyses would have been possible, or inappropriate reporting of table, graph, or meta-analyses)
3= Partially (if appropriate table, graph or meta-analysis and appropriate weights and unit of analysis errors not addressed, but and should have been)
4= Not applicable (e.g. no studies or no data)
5= Can't tell (if unsure)

3.8. a) Does the review report evidence appropriately?

1= The review makes clear which evidence is subject to low risk of bias in assessing causality (attribution of outcomes to intervention), and which is likely to be biased, and does so appropriately
2= Where studies of differing risk of bias are included, results are reported and analyzed separately by risk of bias status

3.8.b) Does the review report evidence appropriately?

1= Yes (Both criteria should be fulfilled (where applicable))
2= No (Criteria not fulfilled)
3= Partially (Only one criterion fulfilled, or when there is limited reporting of quality appraisal (the latter applies only when inclusion criteria for study design are appropriate))
4= Not applicable (No included studies)

3.9. Were factors that the review authors considered as likely explanatory factors clearly described?

1= Yes
2= No

3.10. Was a sensible method used to explore the extent to which key factors explained heterogeneity?

1= Descriptive/textual
2= Graphical
3= Meta-analysis by sub-groups
4= Meta-regression
5= Other

3.11. Did the review examine the extent to which specific factors might explain differences in the results of the included studies?

1= Yes (Explanatory factors clearly described and appropriate methods used to explore heterogeneity)
2= Partially (Explanatory factors described but for meta-analyses, sub-group analysis or meta-regression not reported, when they should have been)
3= No (No description or analysis of likely explanatory factors)
4= Not applicable (e.g., too few studies, no important differences in the results of the included studies, or the included studies were so dissimilar that it would not make sense to explore heterogeneity of the results)

3.12. Comments

Section 4 of 4 Overall assessment of the reliability of the review

4.1. Are there any other aspects of the review not mentioned before which lead you to question the results?

1= Additional methodological concerns - only one person reviewing
2= Robustness

3= Interpretation
4= Conflicts of interest (of the review authors or for included studies)
5= Other
6= No other quality issues identified

4.2. Are there any mitigating factors which should be taken into account in determining the reviews reliability?

1= Limitation acknowledged
2= No strong policy conclusions drawn (including in abstract/ summary)
3= Any other factors

4.3. Comments to specify if relevant, to flag uncertainty or need for discussion

4.4. Based on the above assessments of the methods please provide a summary of the quality of the review. Strengths and limitations should be summarized above, based on what was noted in Sections 2, 3, and 4.