Experimental Agriculture (2023), vol. 59, e9, 1–19
doi:10.1017/S0014479723000054

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

Kenyan farmers appreciate the higher yield of 50% non-pollen producing Maize (*Zea mays*) hybrids

Hugo De Groote[1], Michael K. Ndegwa[1,2], Nancy Muriithi[3], Bernard G. Munyua[1], Sarah Collinson[4] and Michael S. Olsen[1]

[1]International Maize and Wheat Improvement Centre (CIMMYT), P.O. Box 1041-00621, Nairobi, Kenya, [2]Previously: Natural Resources Institute (NRI), University of Greenwich, Medway Campus, Central Avenue Chatham Maritime, Kent ME4 4TB, UK, [3]Kenya Agricultural and Livestock Research Organization (KALRO), Embu, Kenya and [4]Corteva Agriscience, 18369 County Rd. 96, Woodland, CA 95695, USA
Corresponding author: Hugo De Groote; Email: H.Degroote@cgiar.org

(Received 30 September 2022; revised 06 April 2023; accepted 13 April 2023)

Summary
Incorporating the dominant male sterile gene, Ms44, in new maize varieties results in 50% non-pollen producing (FNP) varieties. This makes the varieties more nitrogen efficient and increases yield directly by an average of 200 kg ha^{-1} across yield levels. However, as half of the plants do not shed pollen, the presence of Ms44 in an FNP variety is clearly visible. This technology can improve food production and security in the African maize-based agri-food systems, but only if accepted by farmers. Farmers were therefore invited to 11 on-farm, researcher managed trial sites of FNP varieties in Kenya over 2 years. They were asked to identify the traits they find important in evaluating maize varieties and to score the FNP varieties, as well as their conventional counterparts, on these criteria (including yield, resistance to pests, and cob size) and overall, using a five-point hedonic scale. In total, 2,697 farmers participated, of which 62% were women. Farmers mentioned many traits they find important, especially yield and related traits, early maturity, and drought resistance, but also tassel and pollen formation. In 2017, mid-season, participants scored FNP varieties lower than conventional varieties on tassel and pollen formation, indicating that farmers could distinguish the trait. FNP varieties still received higher scores for yield and overall evaluation. In mid-season 2018, participants no longer scored FNP varieties lower for pollen formation as they now understood the technology. In both years, at the end-season evaluation, scores for tassel formation were not different, but participants scored FNP varieties higher for yield and overall. We conclude that farmers recognized the FNP trait but did not mind it as they clearly favored its yield advantage. The FNP technology, therefore, has high potential not only to increase maize yields, food production, and food security in the agricultural systems of Africa but also to increase varietal turnover and the adoption of new, high-yielding, climate-smart maize hybrids.

Keywords: Africa; Farmer evaluation; Maize; Male sterility; Participatory variety evaluation

Introduction
Sub-Saharan Africa (SSA) has difficulties feeding its population. The major agricultural systems are based on maize as the primary food crop, but its production has not kept up with the rapidly increasing population. In Kenya, the population increased from 8.4 to 44.9 million people between 1961 and 2014 (an increase of 430%); during the same period, yield increased only from 1.4 to 1.8 kg ha^{-1} (43%) and stagnated in the last decades (FAOSTAT, 2022). Consequently, maize production per capita has decreased substantially, from 145 kg/person in the 1970s to only

77 kg/person between 2004 and 2014 (FAOSTAT, 2020). In the Kenyan maize-based agricultural system, farmers have adopted mainly improved maize varieties, but they are not as enthusiastic about fertilizer, and the quantities applied are too low to restore soil fertility (De Groote *et al.*, 2005; Duflo *et al.*, 2008; Jena *et al.*, 2020). In most of SSA, fallow areas have disappeared, and the use of organic and chemical fertilizers is too low to maintain soil fertility (Binswanger-Mkhize and Savastano, 2017). Nitrogen is one of the most crucial crop nutrients (together with phosphate and potassium) and its incomplete application in fertilizer is one of the major limiting factors for maize production in Africa (Vanlauwe *et al.*, 2011). Therefore, developing maize varieties that are more nitrogen efficient can make a significant contribution to increasing yields and improving food security in Kenya and SSA.

A new hybrid seed production technology (SPT) incorporates the dominant mutant male-sterility gene Ms44 in the female parent (Fox *et al.*, 2017). The technology produces cleaner seed, as pollination from the female parent is not possible, leading to lower seed production cost as detasseling is not needed. When using this hybrid seed, half the plants carry the MS44 allele and their male flowers are sterile. Early in plant development, if nitrogen is limited, the nitrogen normally used for tassel and pollen development will be used for ear and kernel growth instead (Loussaert *et al.*, 2017). As a result, the technology increases yields under various environmental conditions, especially nitrogen stress (Collinson *et al.*, 2022; Fox *et al.*, 2017).

As half of the plants do not shed pollen, the presence of Ms44 in an FNP variety is clearly visible; further, the non-pollen producing (FNP) plants show a slight reduction in plant height, tassel size, have thin anthers, and do not contain viable pollen (Supplementary Material Fig. S1). As farmers are keen observers of their crops, they will likely observe the differences in the new FNP varieties. Further, farmers tend to be conservative and risk-averse (Visser *et al.*, 2020) and might object to the atypical varieties, preferring to stick to their old, conventional ones. Therefore, before bringing FNP varieties with the new traits into the farming system and on the market, it is essential to elicit farmers' opinion on the acceptability of their phenotype and to have farmers evaluate the new varieties and compare them to the conventional varieties.

Therefore, this study's specific objectives were (1) analysis of the traits that farmers find important in evaluating maize varieties, in particular the importance of tassel formation and pollen shed and (2) farmer evaluation of the new FNP varieties on those criteria, in comparison to conventional varieties. Farmers were invited to different trial sites of FNP varieties in Kenya. They were asked to identify the traits they find important in evaluating maize varieties and to quantify the importance of each of these criteria. Further, we asked farmers to score the FNP varieties as well as their conventional counterparts (identified only by a number) on these criteria (including yield, resistance, and cob size) and overall. To understand farmers' reaction to the FNP trait, we had to include tassel formation and pollen shed in the evaluation. However, drawing attention to these might negatively affect the assessment of other traits of the FNP varieties or their overall evaluation. To analyze this possible bias, only half the farmers, randomly selected, were asked to evaluate the varieties on tassel formation and pollen shed, while those criteria were skipped for the other participants, to avoid drawing attention to what might be perceived as negative.

The results of this study bring important and timely information on the FNP technology that has high potential to increase maize yields, food production, and food security in the agri-food systems of SSA. The potential of the technology stems from increasing nitrogen efficiency in a continent where insufficient nitrogen application remains a major constraint to yield. Moreover, as the maize-based agricultural systems in SSA lag behind those from other regions in varietal turnover (Abate *et al.*, 2017; Walker *et al.*, 2015), the technology could also be beneficial to the broader system by increasing the adoption of new, high-yielding, climate-smart maize hybrids.

Materials and Methods

Conceptual framework and overview

We base our conceptual framework on Lancaster's model of consumer choice, which proposes that consumers derive satisfaction not from the goods themselves but from the attributes they provide (Lancaster, 1966). Similarly, farmers derive pleasure from the different qualities of new maize varieties and compare them to the characteristics of the varieties they currently use. Farmers therefore evaluate maize varieties on a range of traits (or attributes) that differ in importance (yield is typically more important than processing quality) (De Groote *et al.*, 2002). In past experiences, tassel development and the quantity of pollen shed were not usually mentioned as important criteria by farmers in the selection of their varieties (De Groote *et al.*, 2004, 2002; Siambi *et al.*, 2002), at least not in open-ended questions.

To make sure that we understand the importance of the tassel and pollen in farmers' evaluation of maize varieties, these traits need to be specifically asked about. At the same time, by raising the question, the researcher draws attention to a trait that might otherwise go unnoticed, increasing its importance and possibly biasing the evaluation. This can negatively affect farmers' evaluation of the varieties with a low score on that criterion and so affect the overall score. The effect of including these two criteria in the overall evaluation can, however, be measured by appropriately adjusting the experimental design. We therefore assigned the participating farmers randomly to two treatment groups, and in the first one including the evaluation of the tassel (treatment group 1), and in the second one including the evaluation of both tassel and pollen (treatment group 2), and comparing them to a control group where these criteria were not included. At the end-season evaluation, as the pollen was no longer visible, the pollen criterion was dropped from the evaluation, so treatment groups 1 and 2 were merged. Participants in the control group were not asked to evaluate tassel or pollen. Because in the first year (2017) mid-season, the evaluations of the varieties by the two treatment groups were very similar, the groups were also merged during the mid-season evaluation in the second year (2018).

Study design/empirical framework

The evaluation was superimposed on a set of trials with FNP varieties conducted in Kenya's farmers' fields in the long rainy seasons of 2017 and 2018 (Collinson *et al.*, 2022). The trial compared conventional hybrids with the FNP version of the same hybrid, with two replicates in each site. In 2017, the hybrids used were combinations of two female lines (KTN71-6 and CML550) and two male lines (CML572 and CML494), resulting in four entries (Table 1). We repeated each entry once, resulting in 16 plots in each trial. In this first year, the varieties were chosen based more on availability than suitability to the regions, and the varieties were all single crosses. In the second year, 2018, more varieties were available. Eight varieties, of which four were three-way crosses, were included in the trial, with more varieties adapted to the study areas.

Previous studies with farmers in participatory variety evaluation during participatory rural appraisals provided a set of traits that farmers found necessary in evaluating maize varieties in Kenya (Bett *et al.*, 2002; De Groote *et al.*, 2002; Siambi *et al.*, 2002). In 2017, we included 13 of these traits (rows 1–13 in Table 2). Based on discussions with the farmers, we included four more criteria in 2018 (rows 14–17). To confirm the importance of these criteria to the participants of this study, we asked them to give the criteria a score for importance (0 = not important, 1 = somewhat important, 2 = important, 3 = very important) (Table 2).

Participants were asked to evaluate the different entries on these criteria. In 2017, they assessed the four entries and their replicates, with 16 plots in total. In 2018, there were twice as many entries, 16, so it became difficult for one participant to evaluate them twice. Therefore, each participant was only asked to consider one of the two replicates, randomly assigned. To score the entries, participants used a 5-point Likert scale (Likert, 1932). Experience has shown that a 5-point

https://doi.org/10.1017/S0014479723000054 Published online by Cambridge University Press

Table 1. Varieties used in the trials, each variety was evaluated with and without Ms44, and with two reps. in each trial

Year	Symbol	Pedigree
2017	V1	KTN71–6/CML572
	V2	KTN71–6/CML494
	V3	CML550/CML572
	V4	CML550 /CML494
2018	V1	CML395/CML572
	V2	CML395/CML540
	V3	CML550/CML572
	V4	LaPostaSeqC7/CML444
	V5	CML550/CML494//CML536
	V6	CML550/CML572//CML444
	V7	KTN71–6/CML494//CML572
	V8	KTN71–6/CML572//CML539

Note: '/' indicates single cross, and the first variety is the female parent; '//' indicates the second cross, so A/B//C is a three-way hybrid where A is the female of the single-cross A/B, and A/B is the single-cross female parent of three-way hybrid A/B//C.

Table 2. Tool for evaluations with the traits evaluated, and experimental treatments among participants

Criteria no.	Treatment group	Year	Traits	**How important is this trait?**[a]	**For each plot, please evaluate the variety for this trait on a scale of A to E**[b]		
					Plot 1	Plot 2	. . .
1	All	2017 and 2018	Germination/Crop stand				
2	All	2017 and 2018	Height				
3	All	2017 and 2018	Stalk thickness				
4	All	2017 and 2018	Number of cobs per plant				
5	All	2017 and 2018	Cob size				
6	All	2017 and 2018	Barrenness level				
7	All	2017 and 2018	Yield				
8	All	2017 and 2018	Biomass (for fodder)				
9	All	2017 and 2018	Resistance to stalk borer				
10	All	2017 and 2018	Drought resistance				
11	All	2017 and 2018	Foliar disease resistant				
12	All	2017 and 2018	Tillers development				
13	All	2017 and 2018	Early maturing				
14	All	2018 only[c]	Husk cover				
15	All	2018 only	Drooping of the ear				
16	All	2018 only	Cob rot resistance				
17	All	2018 only	Lodging resistance				
14	Treatment 1 and 2[d]	2017 and 2018	Good tassel formation				
15	Treatment 2 only	2017 and 2018	Amount of pollen shed (good pollination)				
20	All	2017 and 2018	Overall evaluation (note: not an average)				

[a]Codes: 0 = not important, 1 = somewhat important, 2 = important, 3 = very important.
[b]Codes: A = like very much, B = like, C = neither like nor dislike, D = dislike, E = dislike very much.
[c]In 2108, four criteria were added.
[d]In the mid-season of 2017, farmers were randomly assigned to three groups: control, treatment 1, and treatment 2; all participants evaluated the varieties on criteria 1 to 13 and 16, both treatment groups also on criterion 14, only those in treatment 2 on criterion 15. In 2018, treatment groups 1 and 2 were merged.

scale is convenient for farmers and can easily be translated into any language (De Groote *et al.*, 2002, 2010). However, experience has also shown that using numbers for the scores can be confusing, as "1" can indicate both a very good or a very poor score. The numbers were therefore replaced by letter scores, which corresponded to the Kenyan school system (Worku *et al.*,

https://doi.org/10.1017/S0014479723000054 Published online by Cambridge University Press

2020). The options were A (like very much), B (like), C (neither like nor dislike), D (dislike), and E (dislike very much). In 2017, farmers were randomly assigned to different groups: control (participants in this group were asked to evaluate the entries on 13 specific traits and overall, but not on tassel formation or pollen shed); treatment 1 (participants evaluated entries on the same traits plus the trait “good tassel formation”); and treatment 2 (participants evaluated entries on the same traits as those in treatment 1 plus the trait “amount of pollen shed”). As the results of 2017 indicated that the results of treatments 1 and 2 were very similar, these two treatment groups were merged into one group in 2018, and the members evaluated the entries on both tassel and pollen. All criteria were expressed on the questionnaire in both English and Kiswahili, the two national languages in Kenya. Depending on the situation, the criteria were translated into local languages in the different counties. The questionnaires are found in the Supplementary Materials (S2 and S3).

In selecting varieties, farmers are also concerned about other factors like seed source, price, availability, and the possibility of using their own saved seeds. However, in this study, we were only comparing conventional hybrids to the same varieties in which the FNP trait was added. In this experimental and study design, therefore, all other factors remain constant.

Power calculations indicated that about 120–166 farmers were needed to make a distinction between the two types of varieties. However, for a design with three treatment arms, we needed at least 200–250 participants, spread over several sites. Finally, to let farmers observe tassels and pollen shed, the evaluations needed to be conducted at the mid-season and, to observe yield, also at the end-season.

Data collection and participants

In 2017, participatory evaluations were organized in five trial sites at the mid-season and eight at the end-season. In 2018, the breeders had dropped two of the eight trial sites for different reasons and replaced with other nearby trial sites, and evaluations took place in all these eight sites at both mid- and end-seasons. The sites were located in six counties in both western and eastern Kenya (Table 3, Supplementary Material Fig. S4). The mid-season evaluations took place from June to early July, and the end-season evaluations took place from the end of July to August, on one day per site. Altogether 1,006 people participated in 2017, with more (73%) at the end-season and 1,691 in 2018 (about half in each season). A majority of participants (62%) were women, and this was consistent over seasons, sites, and years, with few exceptions.

The evaluations followed the standard procedure. First, scientists and farmers were introduced to each other. Then the methods and procedures were explained, and each trait was explained and discussed in detail. The scores to be used for the importance of the traits as well as those used for evaluation were explained. Afterward, participants were organized into groups of 5–10, each accompanied by an enumerator. The participants first filled in an informed consent form, followed by a short socioeconomic questionnaire that included questions on age, sex, education, income, and so forth (Supplementary Materials S2 and S3). They then proceeded to the field where they evaluated the different plots—16 plots in each trial, numbered consecutively, without any indication or mention of variety or FNP—on the different traits. After the participatory variety evaluation, participants regrouped and were offered a snack and a drink. At this point, the principle of the Ms44 gene and the FNP trait was explained to them, what it looked like and how it worked, and they subsequently engaged in an open group discussion, where they could ask more questions and express their opinions.

Data analysis

First, the importance of the different traits was assessed by averaging the importance scores given by participants on a scale of 1 to 4, and compared by season, as farmers can have different priorities over the different seasons. Only participants of the treatment groups were asked to state the

https://doi.org/10.1017/S0014479723000054 Published online by Cambridge University Press

Table 3. Number of participants, by site and gender, in the farmer evaluations of Ms44

				Mid-season				End-season				Overall		
				Number participants				***N* participants**				***N* participants**		
Year	Site #	County	Site	Women	Men	Total	Date	W	M	T	Date	W	M	T
2017	3	Kirinyaga	Gichugu1					18	10	28	28-Aug	18	10	28
	4	Kirinyaga	Gichugu2	47	26	73	07-Jul	31	11	42	29-Aug	78	37	115
	5	Kirinyaga	Mwea East	26	16	42	06-Jul	16	6	22	31-Aug	42	22	64
	13	Embu	Runyenjes	13	4	17	05-Jul	66	18	84	30-Aug	79	22	101
	18	Kakamega	Ileho					77	42	119	25-Aug	77	42	119
	19	Bungoma	Kabula	46	54	100	12-Jul	69	94	163	23-Aug	115	148	263
	25	Kakamega	Shinyalu					77	46	123	24-Aug	77	46	123
	29	Busia	Nambale	18	19	37	13-Jul	77	79	156	22-Aug	95	98	193
		Total		150	119	269		431	306	737		581	425	1006
		%		56	44	100		58	42	100		58	42	100
2018	2	Embu	Karurumo A[a]	44	32	76	12-Jun	64	33	97	26-Jul	108	65	173
	4	Embu	Karurumo B	107	32	139	11-Jun	70	17	87	26-Jul	177	49	226
	7	Kirinyaga	Togonye	17	10	27	14-Jun	79	34	113	25-Jul	96	44	140
	9	Kirinyaga	Mururi[a]	107	53	160	13-Jun	82	29	111	25-Jul	189	82	271
	12	Kakamega	Lubao	53	26	79	04-Jul	79	34	113	01-Aug	132	60	192
	13	Kakamega	Ilala	63	46	109	03-Jul	67	46	113	31-Jul	130	92	222
	19	Bungoma	Kabula	48	64	112	03-Jul	57	54	111	31-Jul	105	118	223
	20	Kakamega	Murhanda	63	57	120	04-Jul	86	38	124	01-Aug	149	95	244
			Total	502	320	822		584	285	869		1086	605	1691
			%	61	39	100		67	33	100		64	36	100

importance of the FNP-related traits: pollen produced and tassel formation. However, the importance of these traits was asked for after providing information on the other traits and before observing the varieties in the field visits. Therefore, the treatment could not affect the stated importance of other traits.

Next, for each trait, the ordinal scores were converted from their alphabetical form to numerical values ($A = 5$ to $E = 1$), and the average scores calculated for all the plots, to compare the FNP entries with their conventional counterparts graphically and statistically using pairwise t-tests. However, even though the variables now have numerical values, they are still on an ordinal scale, not an interval scale. This means that a value of four is higher than a value of two, but it does not mean that a value of four is twice that of a value of two, as it would be for an interval scale. Most statistical calculations, such as mean values and standard errors, assume interval scales and their use with ordinal scales is problematic. Therefore, specific analytical methods for this type of data have been developed, in particular ordinal regression (McCullagh, 1980), which does not require the assumptions of the interval scale. Ordinal regression has been specifically recommended for the analysis of farmer evaluation scores (Coe, 2002). It has been used in East Africa for different technologies including integrated pest and soil fertility management (De Groote *et al.*, 2010) and farmer participatory evaluation of new maize varieties (Worku *et al.*, 2020). For comparing varieties, participants typically evaluate different varieties simultaneously, and those scores are likely to be correlated. Hence, random effects should be added to the model, as was done in a previous study on farmers' evaluation of maize varieties in East Africa (Worku *et al.*, 2020). The ordinal model with random effects was used for this study to analyze farmers' evaluations and compare FNP vs. conventional varieties for the different traits. We also analyzed the effect of the different treatments by including them in the model.

Finally, to analyze the importance of the different criteria in the actual evaluation, we regressed the scores that the varieties received for individual traits on the overall scores that they received. The coefficients of the individual traits could then be interpreted as their weights in the overall score, as the 'revealed importance', and compared to the importance scores given to the criteria earlier, which could be interpreted as the 'stated importance'. Unlike that of the stated importance,

https://doi.org/10.1017/S0014479723000054 Published online by Cambridge University Press

Table 4. Descriptive statistics of the participating farmers

Group	Characteristic	2017 ($N = 903$)		2018 ($N = 1691$)		Pooled means ($N = 2594$)	
		Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Personal	Gender (1 = female, 0 = male)	0.58	0.49	0.64	0.48	0.62	0.36
	Age	44.06	14.83	44.88	15.38	44.59	11.28
	Farming experience	17.48	13.20	17.23	13.78	17.32	10.09
	Formal education years	8.33	5.00	7.72	3.73	7.93	2.99
	Total income (previous year, KES)	92 617	223 475	78 320	121 420	83 297	110 982
Farm	Own farm size (ha)	0.85	1.09	0.7325	0.744638	0.77	0.62
	Area under maize (ha)	0.49	1.24	0.44921	0.400648	0.46	0.51
	Purchased improved maize seed	0.86	0.35	0.92	0.28	0.90	0.22
	Purchased fertilizer	0.86	0.35	0.96	0.19	0.93	0.17
	Cattle owned (number)	1.71	2.08			1.71	2.08
	Oxen owned (number)	0.39	0.82			0.39	0.82

analysis of the revealed importance also allows us to compare the treatment groups with the control to see if drawing attention to the tassel and pollen affected the overall scores or the scores on other traits.

Results

Characteristics of participating farmers

Participants were adults of all ages (from 17 to 88), with a majority of women (62%) (Table 4). Most participants were farmers with many years of farming experience, on average 17 years. Most had also finished primary education, with an average of 8 years of formal education (the length of primary school education in Kenya). Average annual cash income was KES 83 297 (about US$ 830), about half of which came from agriculture (a bit more from crops than from livestock), a quarter from salary and a quarter from business and trade. Most participants owned their farm, with an average size of almost one ha (0.77 ha), about half of which in maize (0.46 ha). Nearly all participating farmers used improved maize varieties (90%) and fertilizer (93%). Most participants practiced a mixed cropping/livestock system; about two-thirds of participants owned cattle and a quarter owned oxen.

Farmers' evaluation of maize varieties: stated importance of traits

When participants were asked to score the importance of the different traits (or criteria) on a scale of 0 (not important) to 3 (very important), they gave high scores to most of the traits (Figure 1). During the mid-season evaluation, the traits which received the highest scores on importance were yield, early maturity, cob size, and a number of cobs; these all received an average score of between 2.5 and 2.7. When farmers were asked at the mid-season evaluation if tassel formation was important, they scored the trait very highly (2.68) second only to yield (2.69) (out of a maximum of 3). Similarly, the amount of pollen shed received an importance score of 2.6. Participants were not asked to score all traits for importance during the mid-season evaluation; ear-related traits in particular were not discussed as they could not yet be observed at this point in time. In the end-season evaluation, however, tassel formation only scored 2.44, which put it in the third-to-last position in terms of importance among criteria. Also, most traits in the end-season evaluation received even higher importance scores than in the mid-season evaluation, and almost all criteria received a score between 2.5 and 3. There was no significant difference in the importance of these traits between treatment farmers (those asked to evaluate the tassel and pollen) and control farmers (who were not).

https://doi.org/10.1017/S0014479723000054 Published online by Cambridge University Press

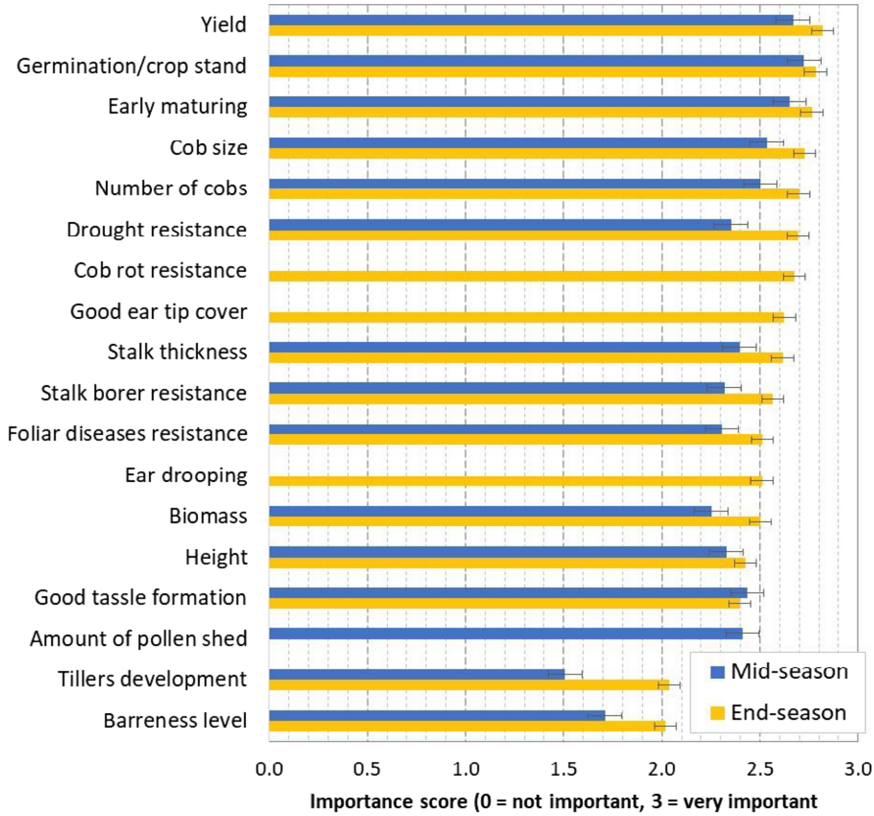

Figure 1. Importance of maize variety evaluation criteria, according to participants (on a scale of 0 = not important to 3 = very important).

Farmer evaluation of FNP varieties in 2017

Farmers were invited to the trials twice, once at the mid-season and once at the end-season, and during each visit they were asked to score the varieties on a 5-point scale for the different traits. By converting the scores to numerical values (from 1 to 5), mean scores could be calculated for the FNP varieties and their conventional counterparts and compared using the pairwise *t*-test. During the mid-season evaluation of 2017, the conventional varieties received significantly higher scores on tassel formation and pollen shed than the FNP varieties, indicating that farmers could visually distinguish FNP from conventional varieties and preferred the latter on these traits (Figure 2). However, in the mid-season, the score for yield and the score for related traits such as cob size were higher for FNP varieties than for conventional ones; otherwise, there was little difference between the scores of both types of variety. Finally, the overall score of the FNP varieties was significantly higher than that of the conventional, pollen-producing varieties.

In the end-season (at harvest), the evaluation for pollen formation was dropped, as there is no pollen to be seen in maize at this stage, but the tassel is still visible, although dried and brown, so farmers in both treatment groups were asked to evaluate this trait. However, the scores for tassel formation were no longer different between FNP and conventional varieties, indicating that participants could no longer tell the difference. This is understandable as, at this stage, the maize plant including the tassel has dried up and turned brown. Further, FNP varieties generally scored better on several individual criteria, especially cob size and yield, and their overall scores were significantly better than those of the conventional varieties. The average scores for all varieties on different traits and seasons are presented in Supplementary Materials Table S5.

https://doi.org/10.1017/S0014479723000054 Published online by Cambridge University Press

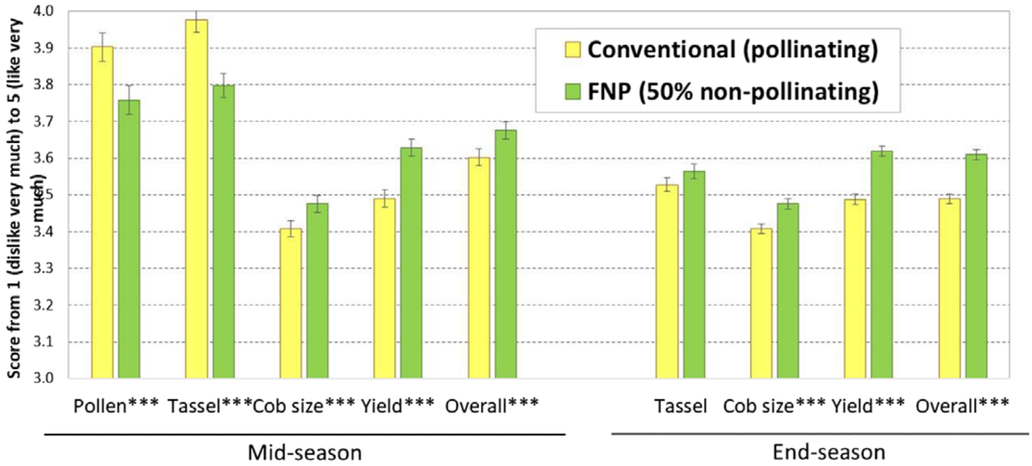

Figure 2. Farmer evaluation of conventional (male fertile) and FNP (50% non-pollinating) varieties in 2017, on a 5-point hedonic scale (1 = dislike very much, 2 = like, 3 = neither like nor dislike, 4 = like, 5 = like very much).

For a more robust statistical analysis, and taking into account the nature of the data, we used an ordinal regression model with random effects. We conducted the analysis for the three major traits relevant to male sterility: tassel formation and pollen shed (expected to be affected negatively), yield (expected to be affected positively), and overall evaluation, the synthesis (Table 5). We also included in the model the four different varieties (the first one being the omitted category), next to the FNP trait. Only farmers in the treatment groups were invited to evaluate the amount of pollen shed (treatment 1) or good tassel formation (treatments 1 and 2).

Treatment 1 participants could clearly distinguish FNP varieties from conventional varieties in the mid-season evaluation by the amount of pollen shed and scored them significantly lower. The coefficient (−0.39) is the log odds ratio and its exponent (0.67) the odds ratio. The odds ratio indicates that the odds of FNP varieties scoring higher than conventional varieties (the probability of FNP varieties scoring higher over the probability that they do not) are 67%. In other words, FNP varieties are less likely to be preferred over conventional varieties on this trait. Farmers in treatment groups 1 and 2 also evaluated the different entries on good tassel formation, and here again FNP varieties scored substantially and significantly lower than their conventional counterparts, with a similar coefficient and odds ratio (67%). Interestingly, there were also differences between varieties, with the second variety scoring significantly better on pollen as well as on tassel. Still during the mid-season evaluation, FNP varieties scored significantly higher than conventional varieties on yield (coefficient of 0.269 or odds ratio of 1.31), indicating that farmers could observe the yield difference that FNP varieties generated and appreciated it. In the overall evaluation, FNP varieties scored better, although the coefficient was small compared to the differences between varieties.

The results of the ordinal regression on the end-season evaluation show that, at this stage, participants did not distinguish between the tassels of FNP varieties and their conventional counterparts. On the other hand, the yield scores for the FNP varieties were substantially and significantly higher than those of the conventional varieties, as in the mid-season evaluation. However, the overall scores for the FNP varieties in the end-season were now substantially and significantly higher than those of their conventional counterparts (coefficient of 0.266 or odds ratio of 1.30).

Finally, we repeated the same ordinal regression from Table 5 on all traits in the study, again with the varieties as co-factors, and sorted the coefficients for FNP vs. conventional varieties (the binary variable in the second row of Table 5). To visualize the results, we then mapped the coefficients to show how the FNP varieties were evaluated differently on these traits (Figure 3) at both the mid-season (Panel A) and the end-season (Panel B). The results show that at the mid-season,

https://doi.org/10.1017/S0014479723000054 Published online by Cambridge University Press

Table 5. Statistical analysis of farmer evaluation of Ms44 varieties in 2017, using ordinal regression with random effects

Variable	Mid-season				End-season		
	Pollen (Tr 2 only)	Tassel (Tr 1 + Tr2)	Yield (all)	Overall (all)	Tassel (Tr1 + Tr2)	Yield (all)	Overall (all)
FNP (1 = FNP variety,	−0.322***	−0.393***	0.269***	0.148**	0.08	0.272***	0.266***
0 = conventional variety)	(0.099)	(0.083)	(0.057)	(0.058)	(0.048)	(0.034)	(0.034)
V2 (KTN71–6/CML494)[a]	0.232	0.183	0.732***	0.683***	0.548***	0.875***	0.932***
	(0.140)	(0.117)	(0.081)	(0.081)	(0.069)	(0.049)	(0.050)
V3 (CML550/CML572)	0.815***	0.583***	1.219***	1.336***	0.064	0.607***	0.546***
	(0.142)	(0.119)	(0.083)	(0.084)	(0.068)	(0.049)	(0.049)
V4 (CML550/CML494)	−0.006	0.113	0.586***	0.534***	−0.3***	0.05	0.068
	(0.139)	(0.116)	(0.080)	(0.081)	(0.068)	(0.049)	(0.048)
σ^2_u	2.062	1.64867	1.158	1.588	1.512	1.288	1.254
N obs.	1,520	2,139	4287	4,288	6,074	11 849	11 802
N part.	101	134	268	268	389	736	735
Wald χ^2	55	49	240	259	59	507	519
Log LH	−1879.04	−2645.765	−5756.87	−5539	−7904	−15584	−15509

[a]The base category is V1 (KTN71-6/CML572).

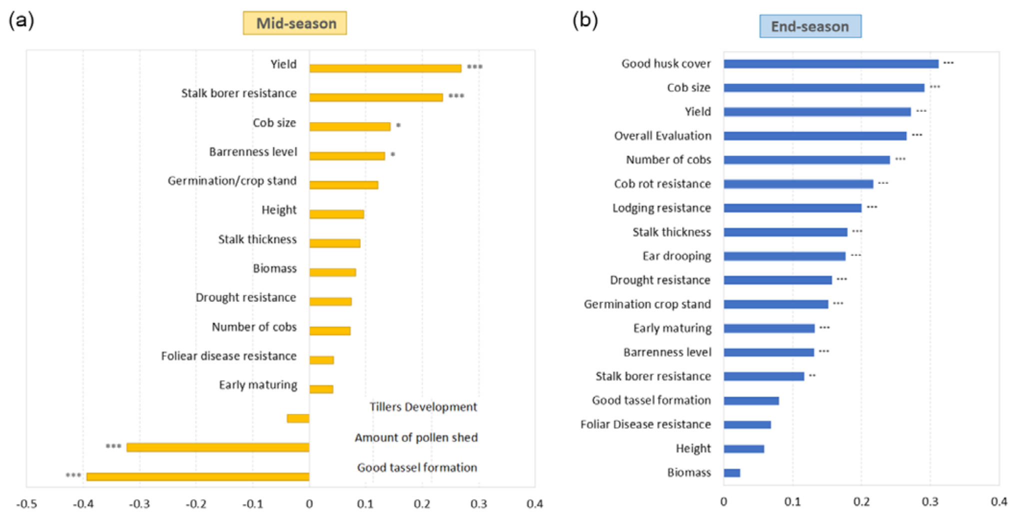

Figure 3. Coefficients of ordinal regression comparing scores of FNP to their conventional counterparts (model includes four varieties and random errors), mid- and end-season 2017.

FNP varieties received better scores on most criteria. Still, the scores were only significantly higher for yield, stalk borer resistance, cob size, and barrenness levels. Still, in the mid-season, the FNP varieties received significantly lower scores for the two relevant traits: tassel and pollen formation. At the end-season evaluation, the FNP varieties scored substantially higher on almost all traits, but not for tassel and pollen formation.

Farmer evaluation 2018

In 2018, the trials were repeated, again at eight sites but with two new ones. This time, eight varieties were included, each with and without FNP, with two replicates. Farmers evaluated all 16 entries, but only one replicate each, with a slightly extended list of traits (Table 2). The results

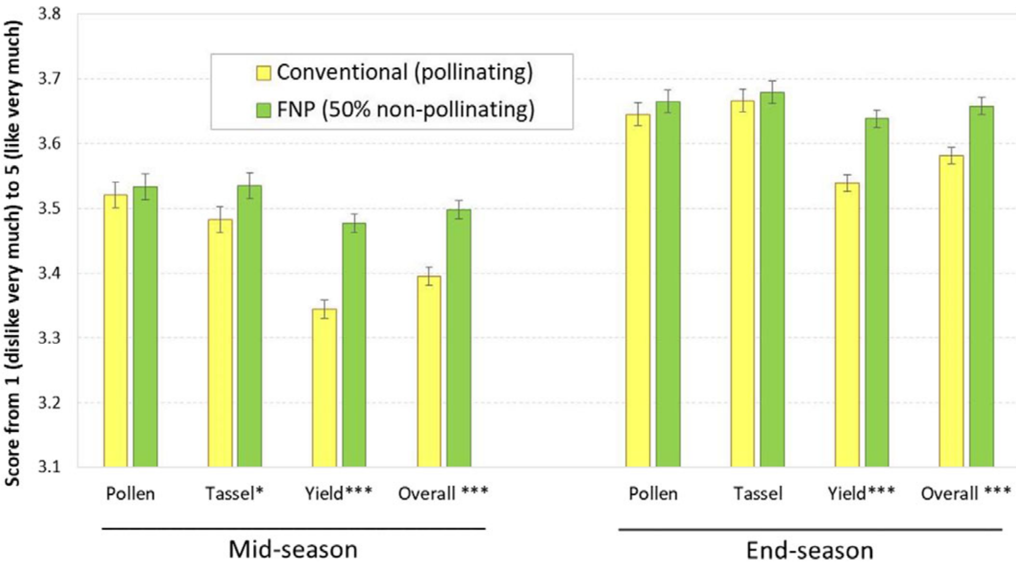

Figure 4. Farmers' evaluation of the conventional (pollen producing) and FNP (50% non-pollinating) entries in 2018 for the key traits.

show that, unlike in 2017, participants in 2018 did not score the FNP entries lower for the amount of pollen shed and even scored them slightly higher for tassel formation (Figure 4). This was against expectations, but during the discussion with farmers at the end of the evaluation sessions, they explained that they remembered from last year that less tassel development and pollen shed were not negative traits. Further, during the same mid-season evaluation of 2018, FNP entries scored markedly higher on yield and overall evaluation compared to conventional varieties. The end-season evaluation of 2018 showed similar results, with no difference in pollen shed or tassel formation scores and significantly higher scores for yield and overall evaluation. The scores for all the traits were, however, much higher compared to the mid-season scores. This might imply that, at the end-season, these features were more developed, and farmers tended to give them higher scores.

For a more robust statistical analysis, we again used the ordinal regression model with random effects (Table 6, first panel). The results confirmed that at the mid-season, the score for pollen was not now different for the FNP varieties, but the score for tassel formation was significant and positive, as were the scores for yield and overall. At the end-season, on the other hand, the scores for tassel and pollen were not significantly different between the FNP and the conventional varieties, but those for yield and overall evaluation were very different (Table 6, second panel).

We repeated the analysis for all the traits in 2018 (Supplementary Material Fig. S6). Except for pollen and tassel, FNP varieties now scored significantly higher than conventional varieties for almost all traits, at both mid- and end-season evaluations. This shows that farmers could tell the difference between FNP varieties and conventional varieties and liked the FNP varieties better, or at least for the traits used in this evaluation.

Effect of the treatments

The study was specifically designed to obtain farmers' opinion on the FNP trait without drawing undue attention to relevant traits. Therefore, only farmers in the treatment groups were asked to evaluate pollen quantity (treatments 1 and 2) and tassel formation (treatment 2). To ascertain if the treatments drew attention to these particular traits, and away from other traits, we analyzed

https://doi.org/10.1017/S0014479723000054 Published online by Cambridge University Press

https://doi.org/10.1017/S0014479723000054 Published online by Cambridge University Press

Table 6. Statistical analysis of farmer evaluation of FNP varieties, using ordinal regression with random effects: mid- and end-season 2018

	Mid-season				End-season			
Variables	Pollen	Tassel	Yield	Overall	Pollen	Tassel	Yield	Overall
FNP	0.034	0.249***	0.190***	0.209***	0.036	0.018	0.190***	0.159***
	(0.05)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)	(0.03)
V2 (CML395/CML540)[a]	−0.367***	−0.429***	−0.633***	−0.613***	−0.701***	−0.709***	−0.633***	−0.809***
	(0.09)	(0.06)	(0.06)	(0.06)	(0.09)	(0.09)	(0.06)	(0.06)
V3 (CML550/CML572)	0.135	0.374***	−0.545***	0.379***	−0.58***	−0.593***	−0.545***	−0.692***
	0.09)	(0.06)	(0.06)	(0.07)	(0.09)	(0.09)	(0.06)	(0.06)
V4 (LaPostaSeqC7/CML444)	−0.639***	−0.741***	−0.836***	−0.952***	−0.978***	−1.015***	−0.836***	−0.877***
	(0.09)	(0.06)	(0.06)	(0.07)	(0.09)	(0.09)	(0.06)	(0.06)
V5 (CML550/CML494//CML536)	−0.255**	−0.214**	0.01	−0.438***	−0.184*	−0.212*	0.01	−0.014
	(0.09)	(0.06)	(0.06)	(0.06)	(0.09)	(0.09)	(0.06)	(0.06)
V6 (CML550/CML572//CML444)	−0.104	−0.252***	−0.31***	−0.352***	−0.564***	−0.674***	−0.31***	−0.408***
	(0.09)	(0.06)	(0.06)	(0.06)	(0.09)	(0.09)	(0.06)	(0.06)
V7 (KTN71-6/CML494//CML572)	0.105	0.453***	−0.066	0.356***	−0.071	0.035	−0.066	0.039
	(0.09)	(0.06)	(0.06)	(0.06)	(0.09)	(0.09)	(0.06)	(0.06)
V8 (KTN71-6/CML572//CML539)	−0.545***	−0.798***	−0.78***	−0.895***	−0.583***	−0.635***	−0.78***	−0.85***
	(0.09)	(0.06)	(0.06)	(0.06)	(0.09)	(0.09)	(0.06)	(0.06)
σ^2_u	1.327	0.881	0.881	1.317	1.111	0.991	0.709	0.851
	(0.12)	(0.11)	(0.06)	(0.08)	(0.10)	(0.09)	(0.05)	(0.05)
N observations	6,443	13 215	13 215	13 119	7,161	7,135	14 220	14 120
N participants	411	828	828	826	474	450	889	886
Wald X^2	136	755	755	870	205	245	479	608
Log likelihood	−8825	−18837	−18837	−18154	−9508	−9431	−19747	−19348
X^2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

[a]The base category is V1 (CML395/CML572).

whether the assignment of participants to treatment or control groups affected their evaluation scores. As illustrated for the overall evaluation, we did not find a significant treatment effect on the scores (Table 7). There was no significant direct effect, meaning that treatment participants do not give different scores in general. There was also no significant cross effect, meaning that treatment participants did not give separate scores for FNP varieties. Analysis for the yield scores produced similar results.

We also analyzed whether, by drawing attention to specific traits, the importance of these traits differed between treatment and control groups. In the stated importance scores, we did not find any differences between the groups. This is understandable, as farmers were asked this at the beginning of the field day, before they had observed the varieties in the field, and the traits for pollen and tassel were put at the end of the list. Therefore, we explored an alternative analysis by regressing the overall score on the scores for different traits (results for 2017 in Figure 5, for 2018 in Table 8). As the coefficients add up to about 1, they can be interpreted as the weight farmers give to the individual traits that are combined into giving an overall score.

The results for the mid-season evaluation of 2017 show that in treatment 2, where participants were asked to score both pollen and tassel, the coefficient for pollen shed was significant. Still, the coefficient for good tassel formation was not (Figure 5). In treatment 1, where farmers were not asked about pollen, the coefficient for tassel formation was significant and, remarkably, almost equal to the coefficient for pollen in treatment 2. This suggests that farmers in treatment 1 were actually judging the varieties by the quantity of pollen shed (or combining the two traits) when evaluating the tassel formation. Adding the evaluation for tassel or pollen did affect the weights of the other traits. Still at the mid-season of 2017, this decreased the weight for yield, the most important trait, and brought the importance of good tassel formation (treatment 1) and amount of pollen shed (treatment 2) to a similar level as yield, as measured by the coefficients, which contrasted strongly with the stated importance (Figure 1). Similarly, at the end-season, the effect of the treatment was that good tassel formation became the second-most-important trait after yield. The effect of the treatment on other traits was, however, not consistent over the seasons.

We repeated the analysis for the 2018 data (Table 8). Again, farmers in the treatment groups evaluated the different varieties for good tassel formation and pollen shed, while those in the control group did not. At the mid-season (Table 8, first panel), the coefficients for both good tassel formation and pollen shed were significant at 1%, but pollen shed was substantially higher (0.32) than tassel formation (0.19). Amount of pollen shed was the third most important trait after yield and early maturing. Adding the evaluations for tassel and pollen did affect the weights of the other traits; it decreased particularly the coefficients for yield, early maturing, cob size, and number of cobs and increased the coefficients for crop stand, stalk thickness, and biomass. During the end-season evaluation also, when farmers were asked to evaluate varieties for the tassel (treatment group), the evaluation score for tassels affected the overall score, substantively as well as significantly (Table 8, second panel). Good tassel formation became the most important trait, overtaking yield which in the control group (and in other studies as well) was always the most important trait. So even though farmers did not distinguish FNP from conventional varieties on tassel formation, the scores they gave to different varieties on this trait affected their overall evaluation.

Finally, the analysis also revealed which criteria were important in farmers' overall evaluation, and how this differed from what they stated as important. Yield consistently received the highest coefficient (on average > 0.40), indicating that it contributed most to the overall evaluation (more than 40%) (Table 8). The only other trait that received a similar value was early maturity, but only at the mid-season evaluation. The next two traits were cob size and the number of cobs, receiving values of around 0.2, and both were strongly related to yield. These were followed by a range of traits receiving between 0.1 and 0.2, including height, resistance to diseases, germination, good tassel formation, and drought resistance. A range of other traits, while still significant, received lower values.

https://doi.org/10.1017/S0014479723000054 Published online by Cambridge University Press

https://doi.org/10.1017/S0014479723000054 Published online by Cambridge University Press

Table 7. Analysis of the effect of the treatments on overall evaluations (ordinal regression model including varieties and random effects)

		2017						2018					
		Mid-season			End-season			Mid-season			End-season		
Group	Variable	Coeff.	St dev.	***P***	Coeff.	St dev.	***P***	Coeff.	St dev.	***P***	Coeff.	St dev.	***P***
FNP	FNP	0.269	0.101	0.008	0.279	0.052	0.000	0.190	0.045	0.000	0.187	0.044	0.000
Treatment	Treatment	0.220	0.187	0.239	−0.074	0.096	0.445	0.034	0.092	0.711	0.127	0.076	0.093
	FNP × treatment	−0.179	0.123	0.146	−0.023	0.069	0.738	0.038	0.065	0.555	−0.054	0.062	0.385
Variety	V2	0.683	0.081	0.000	0.932	0.050	0.000	−0.613	0.064	0.000	−0.808	0.062	0.000
	V3	1.337	0.084	0.000	0.546	0.049	0.000	0.379	0.066	0.000	−0.692	0.063	0.000
	V4	0.534	0.081	0.000	0.068	0.048	0.163	−0.952	0.065	0.000	−0.877	0.063	0.000
	V5							−0.438	0.064	0.000	−0.014	0.062	0.825
	V6							−0.352	0.064	0.000	−0.408	0.062	0.000
	V7							0.356	0.064	0.000	0.039	0.062	0.527
	V8							−0.895	0.065	0.000	−0.850	0.062	0.000
Model	/sigma2_u	1.549	0.167		1.253	0.082		1.316	0.082		0.848	0.054	
	N obs.	4,288			11 802			13 119			14 120		
	N participants	268			735			826			886		
	Log likelihood	- 5,536			-15 508			-18 154			−19347		
	Wald $chi^2(6)$	266			520			870			610		
	Prob > chi^2	0			0			0			0		

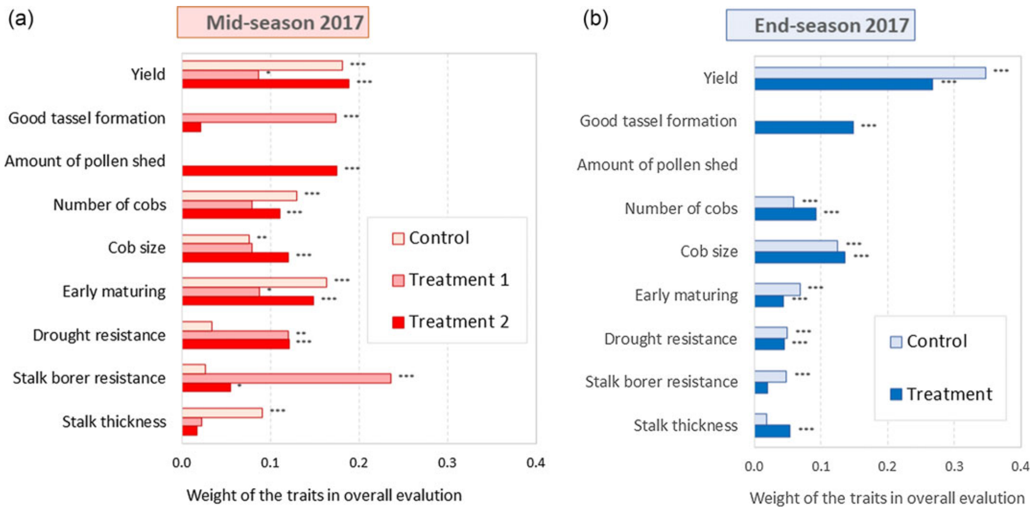

Figure 5. Decomposition of the overall evaluation score over individual evaluation scores, mid-season and end-season 2017 and mid-season 2018, using ordinal linear regression. For the mid-season, farmers in Treatment 1 were asked to evaluate the varieties for the same traits as the control farmers plus the trait 'good tassel formation'; farmers in Treatment 2 evaluated varieties on the same traits as those in Treatment 1 plus the trait 'amount of pollen shed'). In the end-season, as pollen shed could no longer be observed, the two treatment groups were merged, and treatment farmers evaluated the varieties for the same traits as control farmers plus the traits 'good tassel formation'.

Discussion

This paper presents the analysis of an evaluation by farmers of FNP varieties compared with their conventional counterparts at the mid-season and the end-season over 2 years, using a 5-point Likert scale on a range of traits. The results show that when farmers are asked to evaluate maize varieties, they have a wide range of traits or criteria that they use, including tassel formation and pollen shed. Yield-related traits stand out as the most important, including yield, crop stand, cob size, and the number of cobs. Early maturity and drought resistance are also important. This was consistent at both mid- and end-season evaluations and in both years.

At the mid-season evaluation in the first year, FNP varieties scored lower than their conventional counterparts for pollen shed and tassel formation, but higher for yield, cob size, and overall evaluation. This clearly showed that farmers could distinguish the FNP trait at the mid-season evaluation, but they still preferred the FNP varieties over the others because of higher scores on other criteria, including yield. At the end-season evaluation, however, tassel formation in FNP varieties no longer received a lower score, indicating that the trait was no longer observable at harvest. In the second year, at the mid-season evaluation, there was no significant difference for pollen shed and only a small one for tassel formation, for which FNP varieties now received a higher score. This indicated that farmers now understood the technology and its positive effects; this was also confirmed in the discussions with farmers afterwards. In the second year, FNP varieties generally received higher overall scores at both mid-season and end-season, clearly linked to higher scores for yield and related traits such as cob size, and a better overall outlook. This agrees with past studies that reported increased yield in hybrids segregating for male sterility by the use of the Ms44 gene (Fox *et al.*, 2017; Loussaert *et al.*, 2017). The results can be taken as an early indicator of acceptability, as long as the inclusion of the FNP trait leads to higher yields. As a secondary result, participants also scored three-way crosses higher than single crosses.

Asking farmers about different criteria and including specific traits such as pollen and tassel were shown here to affect their evaluation, so this type of study needs to be carefully designed and its impacts analyzed, as demonstrated here. Farmers tend to state many traits as very important,

https://doi.org/10.1017/S0014479723000054 Published online by Cambridge University Press

Table 8. Analysis of overall evaluation on individual evaluations, mid- and end-season 2018, using ordinal linear regression

	Mid-season 2018						End-season 2018					
	Treatment			Control			Treatment			Control		
Trait	Coeff.	Std. Error	Sig.	Coeff.	Std. Error	Sig.	Coeff.	Std. Error	Sig.	Coeff.	Std. Error	Sig.
Yield	0.375	0.031	0.000	0.461	0.032	0.000	0.359	0.044	0.000	0.441	0.044	0.000
Early maturing	0.326	0.027	0.000	0.623	0.027	0.000	0.087	0.039	0.026	0.170	0.038	0.000
Cob size	0.311	0.032	0.000	0.317	0.032	0.000	0.197	0.045	0.000	0.258	0.044	0.000
Number of cobs	0.171	0.032	0.000	0.187	0.032	0.000	0.232	0.046	0.000	0.265	0.045	0.000
Height	0.128	0.031	0.000	0.134	0.030	0.000	0.276	0.045	0.000	0.148	0.044	0.001
Disease resistance	0.105	0.027	0.000	0.194	0.028	0.000	0.111	0.040	0.005	0.038	0.038	0.313
Germination/crop stand	0.205	0.028	0.000	0.093	0.028	0.001	0.197	0.042	0.000	0.302	0.041	0.000
Good tassel formation	0.189	0.031	0.000				0.268	0.046	0.000			
Barrenness level	0.128	0.028	0.000	0.163	0.029	0.000	0.102	0.040	0.012	0.073	0.040	0.066
Drought resistance	0.126	0.027	0.000	0.219	0.028	0.000	−0.023	0.042	0.576	0.098	0.039	0.012
Biomass	0.073	0.028	0.010	0.048	0.028	0.083	0.120	0.041	0.003	0.102	0.040	0.010
Stalk borer resistance	0.040	0.028	0.151	0.056	0.028	0.045	0.056	0.041	0.178	0.062	0.040	0.118
Stalk thickness	0.123	0.033	0.000	0.070	0.031	0.023	0.018	0.045	0.697	0.101	0.044	0.021
Tillers development	0.068	0.023	0.003	0.063	0.024	0.008	−0.062	0.038	0.106	−0.025	0.037	0.507
Husk cover							0.082	0.038	0.031	0.212	0.036	0.000
Drooping of the ear							0.217	0.038	0.000	0.106	0.038	0.005
Cob resistance							0.111	0.040	0.005	0.275	0.039	0.000
Lodging resistance							0.044	0.040	0.270	0.239	0.040	0.000
Amount of pollen shed	0.319	0.030	0.000				0.482	0.045	0.000			
R^2	0.558			0.574			0.281			0.296		
N	6546			6680			3753			3839		

while not making much distinction between them. Regressing overall evaluation over evaluations of different individual criteria, however, shows a more nuanced story and indicates which traits really matter. However, including different traits, as seen when including pollen shed and tassel formation in the treatment groups, affects the weight of the other criteria.

The method used here—asking farmers to score unnamed varieties on a range of traits or attributes on a 5-point Likert scale—is a convenient way to evaluate new maize technology, as has previously been shown for evaluating stress-tolerant varieties (Worku *et al.*, 2020) and pest and soil fertility management (De Groote *et al.*, 2010). The method has some limitations, in particular, that the scores are ordinal categorical data that should be analyzed with ordinal regression, and the results (log-odds ratio) are not easy to interpret. However, as we have shown here, the visual representation of average scores in combination with pair-wise *t*-tests gives similar, although not always equal, results that are much easier to understand. Also, the method does not allow to include other factors, in particular price, availability, and possibility of using farmer-saved seed. In the future, when conducting farmer evaluations of FNP varieties under farmer and market conditions, these factors can be included by estimating willingness to pay for seed with choice experiments or experimental auctions (Horna *et al.*, 2007; Marenya *et al.*, 2021).

Our results also show that farmer evaluation of a large number of traits has limited value as the results tend to be highly correlated. Farmers do observe the tassel and pollen issues and are able to observe yield differences. However, when a variety looks good on yield and general appearance, this tends to affect their scores on other traits that are less easily observed. Further research is needed to see if farmers' positive evaluation of yield would also affect their evaluation of other traits that are not positively correlated with yield, such as early maturity for example. It would also be interesting to compare farmers' scores and physical observations of the same traits, to compare farmers' views with the actual crop performance and the breeders' objectives. A final limitation is that the evaluations were done on-farm but in researcher-managed trials, so additional evaluations under farmer conditions would be an important last step in the evaluation of FNP varieties.

Conclusions

We conclude from our results that farmers' appreciation of FNP is positive and will likely not be a hindrance to the development and dissemination of FNP varieties in maize-based agricultural systems. The expected benefit from FNP varieties is ~200 kg ha^{-1}. The production of FNP varieties will be driven by a novel hybrid production process using the dominant male sterile Ms44 gene and a maintainer line. This SPT will enable seed companies to produce high-quality hybrids without the need for detasseling and at a lower cost. Moreover, by incorporating the trait into selected new hybrids, the technology could help to increase the varietal turnover of maize hybrids, which is currently lagging in SSA. Therefore, the combined effects of increased yield and increased varietal turnover from FNP technology are likely to make a substantial impact on food production and food security in the maize-based agri-food systems of SSA. For a proper impact assessment of the technology, monitoring of the adoption of the technology by seed companies, the amount of FNP seed marketed, and the yield increase in farmers' fields remains essential.

Supplementary Materials. Supplementary materials include the questionnaires for 2017 (SM1) and 2018 (SM2), the average scores for different traits in 2017 (SM3), the regression coefficients for the NFP trait for 2018 (SM4), and the decomposition of overall evaluation (regression on overall score on scores for specific traits). The data can be found in the repository and are available online (Collinson *et al.*, 2021).

For supplementary material for this article, please visit https://doi.org/10.1017/S0014479723000054

Acknowledgements. The authors thank the Seed Production Technology for Africa (SPTA) project, the Bill and Melinda Gates Foundation (Investment ID OPP1137722), and the CGIAR Research Program (CRP) MAIZE for their financial and technical support of this study. We thank the staff of KALRO (Embu and Kakamega Centers) and CIMMYT for their technical and logistical support in setting up the trials and organizing the farmer evaluations, in particular Dr. Ligeyo Otieno and Gabriel Migwi. We thank Elizabeth Waygood for copy-editing the manuscript.

Author Contributions. HDG and MO designed the study, MO and SC designed the field trials, MN and NM organized and coordinated the trials in the field and conducted the discussions with the farmers, MN and BM conducted the farmer evaluations and analyzed the data, HDG and MN wrote the first draft of the manuscript, and all authors reviewed and agreed to the final version of the manuscript.

Financial Support. This work was supported, in whole or in part, by the Bill & Melinda Gates Foundation [INV-005460]. Under the grant conditions of the Foundation, a Creative Commons Attribution 4.0 Generic License has already been assigned to the Author Accepted Manuscript version that might arise from this submission.

Conflicts of Interest. SC is an employee of Corteva Agriscience, who owns the rights to the technology. There is no conflict of interest as Corteva Agriscience will provide the technology royalty free to licensed seed companies producing seed for smallholders in the region (https://www.cimmyt.org/content/uploads/2019/03/CIMMYT-SPTA-project-brief-2020-07-web.pdf). The other authors declare no conflict of interest.

References

Abate T., Fisher M., Abdoulaye T., Kassie G.T., Lunduka R., Marenya P. and Asnake W. (2017). Characteristics of maize cultivars in Africa: how modern are they and how many do smallholder farmers grow? *Agriculture and Food Security* **6**, 1–17.

Bett C., De Groote H., Diallo A., Muasya W. and Kiarie N. (2002). Participatory plant breeding for drought resistant Maize varieties in Eastern Kenya. In Kinro F.H., Mukisira E.A., Wamae J.W., Muriithi F.M. and Wasike W. (eds), *Collaborative and Participatory Research for Sustainable Improved Livelihoods. Proceedings of the 7th KARI Biennial Scientific Conference, 13–17 November 2000*. Nairobi, Kenya: Kenya Agricultural Research Institute, pp. 454–458.

Binswanger-Mkhize H.P. and Savastano S. (2017). Agricultural intensification: the status in six African countries. *Food Policy* **67**, 26–40.

Coe R. (2002). Analyzing ranking and rating data from participatory on-farm trials. In Bellon M.R. and Reeves J. (eds), *Quantitative Analysis of Data from Participatory Methods in Plant Breeding*. Mexico, DF: CIMMYT, pp. 46–65.

Collinson S., Hamadzipiri E., De Groote H., Ndegwa M., Cairns J.E., Albertsen M., Ligeyo D., Mashingaidze K. and Olsen M. (2021). *Replication Data for: Incorporating Male Sterility Increases Hybrid maize Yield in Low Input African Farming Systems*, International Maize and Wheat Improvement Centre (CIMMYT) (eds): CIMMYT Research Data & Software Repository Network. Mexico, DF: International Maize and Wheat Improvement Centre (CIMMYT). (https://data.cimmyt.org/dataset.xhtml?persistentId=hdl:11529/10548515)

Collinson S., Hamdziripi E., De Groote H., Ndegwa M., Cairns J.E., Albertsen M., Ligeyo D., Mashingaidze K. and Olsen M.S. (2022). Incorporating male sterility increases hybrid maize yield in low input African farming systems. *Communications Biology* **5**, 729.

De Groote H., Okuro J.O., Bett C., Mose L., Odendo M. and Wekesa E. (2004). Assessing the demand for insect resistant maize varieties in Kenya combining participatory rural appraisal into a geographic information system. In Sperling L., Lancon J. and Loosvelt M. (eds), *Participatory Plant Breeding and Participatory Plant Genetic Resource Enhancement: An Africa-wide Exchange of Experiences. Proceedings of a workshop held in Bouake, Ivory Coast. May 7–10, 2001*. Cali, Colombia: CGIAR Systemwide Program on Participatory Research and Gender Analysis, pp. 148–162.

De Groote H., Owuor G., Doss C., Ouma J., Muhammad L. and Danda K. (2005). The Maize Green revolution in Kenya revisited. *Electronic Journal of Agricultural and Development Economics (eJADE)* **2**, 32–49.

De Groote H., Rutto E., Odhiambo G., Kanampiu F., Khan Z., Coe R. and Vanlauwe B. (2010). Participatory evaluation of integrated pest and soil fertility management options using ordered categorical data analysis. *Agricultural Systems* **103**, 233–244.

De Groote H., Siambi M., Friesen D. and Diallo A. (2002). Identifying farmers' preferences for new maize varieties in Eastern Africa. In Bellon M.R. and Reeves J. (eds), *Quantitative Analysis of Data from Participatory Methods in Plant Breeding*. Mexico, DF: CIMMYT, pp. 82–102.

Duflo E., Kremer M. and Robinson J. (2008). How high are rates of return to fertilizer? Evidence from field experiments in Kenya. *The American Economic Review* **98**, 482–488.

FAOSTAT (2020). *FAOSTAT Food Supply – Crops Primary Equivalent*. Rome: FAO. Available at: http://www.fao.org/faostat/en/#data/CC (accessed 9 June 2020).

FAOSTAT (2022). *FAOSTAT Crops and Livestock Data Base*, Vol. **2022**. Rome: FAO.

Fox T., DeBruin J., Haug Collet K., Trimnell M., Clapp J., Leonard A., Li B., Scolaro E., Collinson S., Glassman K., Miller M., Schussler J., Dolan D., Liu L., Gho C., Albertsen M., Loussaert D. and Shen B. (2017). A single point mutation in Ms44 results in dominant male sterility and improves nitrogen use efficiency in Maize. *Plant Biotechnology Journal* **15**(8), 942–952. https://doi.org/10.1111/pbi.12689.

Horna J.D., Smale M. and Von Oppen M. (2007). Farmer willingness to pay for seed-related information: rice varieties in Nigeria and Benin. *Environment and Development Economics* **12**, 799–825.

Jena P.R., De Groote H., Nayak B.P. and Hittmeyer A. (2020). Evolution of fertiliser use and its impact on Maize productivity in Kenya: evidence from multiple surveys. *Food Security* **13**, 95–111. https://doi.org/10.1007/s12571-020-01105-z.

Lancaster K.J. (1966). A new approach to consumer theory. *Journal of Political Economy* **74**, 132–157.

Likert R. (1932). A technique for the measurement of attitudes. *Archives of Psychology* **22**: 1–55.

Loussaert D., DeBruin J., Pablo San Martin J., Schussler J., Pape R., Clapp J., Mongar N., Fox T., Albertsen M., Trimnell M., Collinson S. and Shen B. (2017). Genetic male sterility (Ms44) increases Maize grain yield. *Crop Science* **57**, 2718–2728.

Marenya P.P., Wanyama R., Alemu S. and Woyengo V. (2021). Trait preference trade-offs among maize farmers in western Kenya. *Heliyon* **7**, e06389.

McCullagh P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society. Series B (Methodological)* **42**, 109–142.

Siambi M., Diallo A.O., De Groote H., Friesen D.K. and Muasya W. (2002). Recent developments in participatory plant breeding for maize in Eastern Africa: experiences from Eastern Kenya. In Nigussie M., Tanner D. and Twumasi-Afriyie S. (eds), *Enhancing the Contribution of Maize to Food Security in Ethiopia: Proceedings of the Second National Maize Workshop of Ethiopia, 12–16 November 2001, Addis Ababa, Ethiopia. Addis Ababa, Ethiopia.* Addis Ababa and Mexico: Ethiopian Agricultural Research Organization (EARO) and International Maize and Wheat Improvement Center (CIMMYT), pp. 223–231.

Vanlauwe B., Kihara J., Chivenge P., Pypers P., Coe R. and Six J. (2011). Agronomic use efficiency of N fertilizer in maize-based systems in sub-Saharan Africa within the context of integrated soil fertility management. *Plant and Soil* **339**, 35–50.

Visser M., Jumare H. and Brick K. (2020). Risk preferences and poverty traps in the uptake of credit and insurance amongst small-scale farmers in South Africa. *Journal of Economic Behavior & Organization* **180**, 826–836.

Walker T., Alwang J., Alene A., Ndjuenga J., Labarta R., Yigezu Y., Diagne A., Andrade R., Andriatsitohaina R.M. and De Groote H. (2015). Varietal adoption, outcomes and impact. Chapter 19 in Walker, T.S. and J. Alwang (Eds.). *Crop Improvement, Adoption and Impact of Improved Varieties in Food Crops in Sub-Saharan Africa*. Wallingford, UK: CABI Publishing. pp. 388–405.

Worku M., De Groote H., Munyua B., Makumbi D., Owino F., Crossa J., Beyene Y., Mugo S., Jumbo M. and Asea G. (2020). On-farm performance and farmers' participatory assessment of new stress-tolerant maize hybrids in Eastern Africa. *Field Crops Research* **246**, 107693.

Cite this article: De Groote H, Ndegwa MK, Muriithi N, Munyua BG, Collinson S, and Olsen MS. Kenyan farmers appreciate the higher yield of 50% non-pollen producing Maize (*Zea mays*) hybrids. *Experimental Agriculture*. https://doi.org/10.1017/S0014479723000054